Part II

Detector Description

Chapter 2

Key Design Considerations for the BTeV Detector

In Part II of the proposal, we present the design of the BTeV detector.

We begin with a discussion of the key "drivers" of the detector design: the physics of B production at the Tevatron and the characteristics of the machine itself. These, taken together with the physics goals outlined in Part I, determine the main requirements for the detector design.

After explaining the requirements of the design, we describe the baseline detector for BTeV, which can achieve our currently stated physics goals, and is designed to have the capability to perform measurements of new interesting final states that may be recognized as important in the future. A schematic of the experiment is shown in Fig. 2.1.

The key design features of BTeV include:

- A dipole located on the IR, which gives BTeV an effective 'two arm' acceptance;

- A precision vertex detector based on planar pixel arrays;

- A detached vertex trigger at Level 1 which makes BTeV efficient for most final states, including purely hadronic modes;

- Excellent particle identification using a Ring Imaging Cherenkov Detector (RICH);

- A high quality $PbWO_4$ electromagnetic calorimeter capable of reconstructing final states with single photons, π^o's, η's or η''s, and identifying electrons;

- Precision tracking using straw tubes and silicon microstrip detectors, which provide excellent momentum and mass resolution;

- Excellent identification of muons using a dedicated detector with the ability to supply a dimuon trigger; and

- A very high speed and high throughput data acquisition system which eliminates the need to tune the experiment to specific final states.
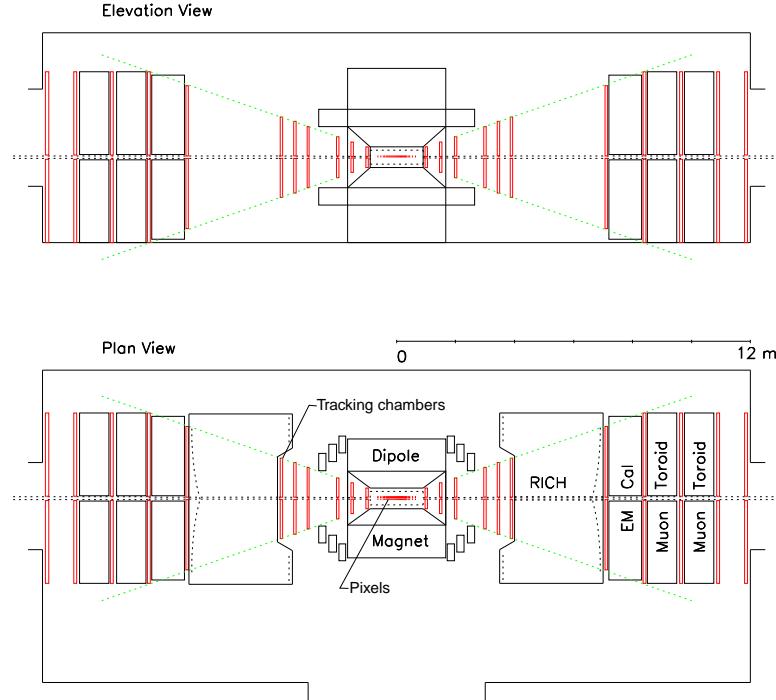
Figure 2.1: Layout of BTeV/C0 Spectrometer

The level of detail in this Part is sufficient to provide the reader with a good overview of the experimental apparatus and a reasonable understanding of the solution to all the various problems associated with carrying out our ambitious program studying B decays. However, many details which an expert might need to understand the detailed implementation and status of development of each technology are not presented here but are placed in appendices, which are referenced in the text.

2.1 Rationale for a Forward Detector

The kinematics of hadronic beauty and charm production play a major role in the design of BTeV. We review the most important features here. In hadron colliders all B species, B^o, B^+, B^o_s, b-baryons, and even B_c mesons, are produced at the same time.

2.1.1 The $b\bar{b}$ Production Cross-Section

It is customary to characterize heavy quark production in hadron collisions with two variables, the momentum transverse to the beams, p_t, and the rapidity,

$$y = \frac{1}{2} \ln \left(\frac{E + p_{\|}}{E - p_{\|}} \right) , \tag{2.1}$$

46

where E is the particle's energy and $p_{\parallel}$ is its longitudinal momentum. Sometimes, the pseudorapidity η is used, where

$$\eta = -ln\left(\tan\left(\theta/2\right)\right) \quad , \tag{2.2}$$

where θ is the angle of the particle with respect to the beam direction. This latter variable was first invented by those who studied high energy cosmic rays, and did not necessarily know the masses of the particles in their detectors.

The $p\bar{p}$ production of b quarks has been measured in the Tevatron at a center-of-mass energy of 1.8 TeV in the central rapidity region $|\eta| < 1$ by CDF [1] and D0 [2], and in the forward region $3.2 > y > 2.4$ by D0 [3]. Both CDF and D0 find that the $b\bar{b}$ production cross-section in the central region is underestimated by the Mangano, Nason and Ridolfi (MNR) next-to-leading order QCD calculation [7] by a factor of approximately two. Since the QCD calculation predicts a cross-section of 50 μb, when integrated over η and p_t, using the data in the central regions leads to a total $b\bar{b}$ production cross-section of 100 μb. The D0 central and forward data are shown in Fig. 2.2.
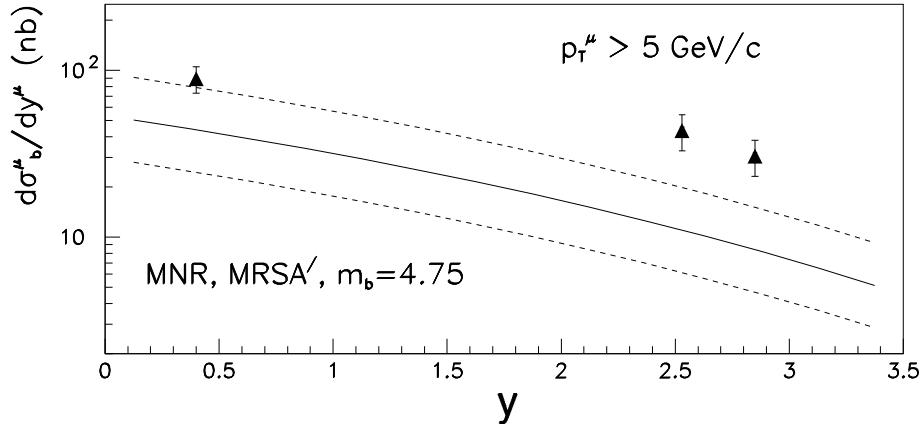

Figure 2.2: The $b\bar{b}$ cross-section as a function of the rapidity of muons from b decay, y^{μ}, measured by D0 for both the forward and central rapidity regions, using muons from b decays with $p_t > 5$ GeV/c. The solid curve is the prediction of the next-to-leading order QCD calculation for a b-quark mass of 4.75 GeV. The dashed curves represent the estimated theoretical 1σ error band.

The measured cross-section in the higher y^{μ} region is 3.6 ± 0.8 times higher than the QCD calculation, leading to a total estimated $b\bar{b}$ production cross-section of 180 μb. BTeV will operate in the range $1.9 > \eta > 4.5$. While we have no reason to dispute the D0 measurement, we will conservatively normalize our estimates to a $b\bar{b}$ production cross-section of 100 μb.

There is some evidence from HERA that the fragmentation of charmed particles is influenced by the leading quarks in the beam so that the fragmentation produces, in some cases, faster D's than the parent c-quarks [4]. This effect is expected to be smaller for b quarks at the center-of-mass Tevatron. If such an effect were present it would increase BTeV's
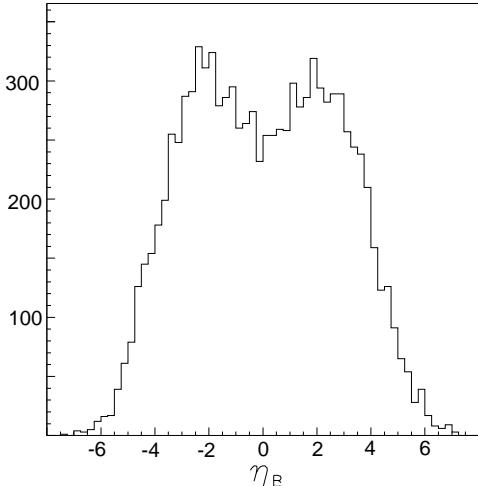
Figure 2.3: The B yield versus η.

acceptance in the forward direction and it would also generate a small asymmetry between B mesons and $\overline{B}$ mesons in the proton direction (opposite in the antiproton direction). All possible production and instrumental asymmetries have to be checked in any case, to obtain accurate measurements of CP violation.

2.1.2 Characteristics of Hadronic b Production

According to QCD calculations of b quark production, the b's are produced "uniformly" in η and have a truncated transverse momentum, p_t, spectrum, characterized by a mean value approximately equal to the B mass [5]. The distribution in η is shown in Fig. 2.3.

There is a strong correlation between the B momentum and η. Shown in Fig. 2.4 is the $\beta\gamma$ of the B hadron versus η from the Monte Carlo physics generator Pythia at $\sqrt{s} = 2$ TeV. It can clearly be seen that near η of zero, $\beta\gamma \approx 1$, while at larger values of $|\eta|$, $\beta\gamma$ can easily reach values of 6. This is important because the observed decay length varies with $\beta\gamma$ and, furthermore, the absolute momenta of the decay products are larger allowing for a suppression of the multiple scattering error.

Since the detector design is somewhat dependent on the Monte Carlo generated b production distributions, it is important to check that the correlations between the b and the $\overline{b}$ are adequately reproduced. Fig. 2.5 shows the azimuthal opening angle distribution between a muon from a b quark decay and the $\overline{b}$ jet as measured by CDF [6] and compares it with the MNR next-to-leading order QCD predictions [7].

The MNR model does a good job representing the shape, which shows a strong back-to-back correlation. The normalization is about a factor of two higher in the data than the theory, which is generally true of CDF b cross-section measurements.

The "flat" η distribution hides an important correlation of $b\overline{b}$ production at hadronic colliders. In Fig. 2.6 the production angle of the hadron containing the b quark is plotted versus

48
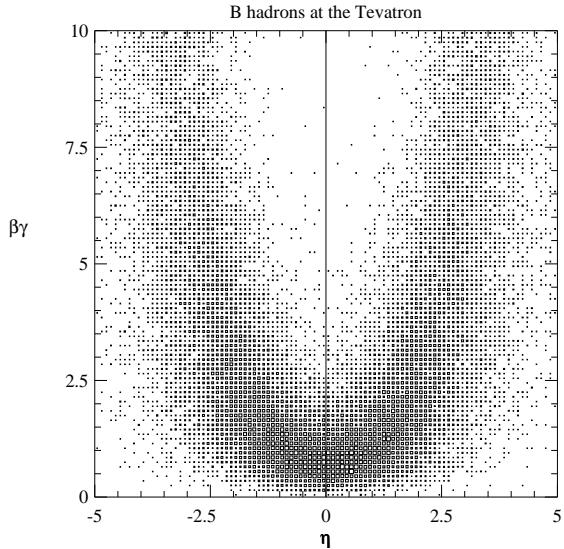
Figure 2.4: $\beta\gamma$ of the B versus η.
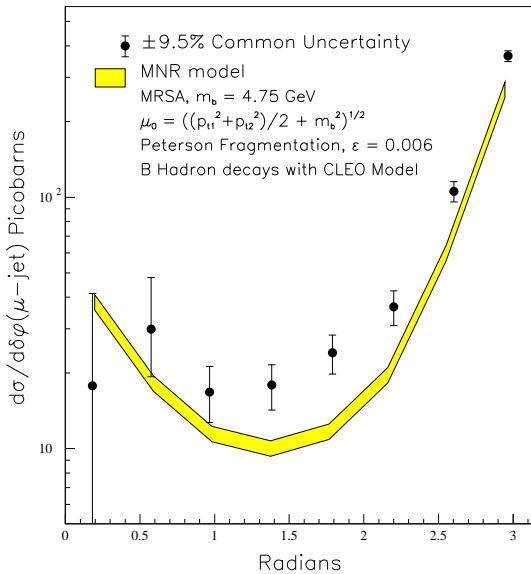
Figure 2.5: The differential $\delta\phi$ cross-sections for $p_T^\mu > 9$ GeV/c, $|\eta^\mu| < 0.6$, $E_T^{\bar{b}} > 10$ GeV, $\left|\eta^{\bar{b}}\right| < 1.5$ compared with theoretical predictions. The data points have a common systematic uncertainty of $\pm 9.5\%$. The uncertainty in the theory curve arises from the error on the muonic branching ratio and the uncertainty in the fragmentation model.

the production angle of the hadron containing the $\bar{b}$ quark. Here zero degrees represents the direction of the incident proton and 180 degrees, the incident anti-proton. There is a very strong correlation in the proton or the anti-proton directions: when the B is forward the $\overline{B}$ is also forward. (We call both the proton and anti-proton directions forward.) This correlation

between B and $\overline{B}$ production is not present in the central region (near 90 degrees).
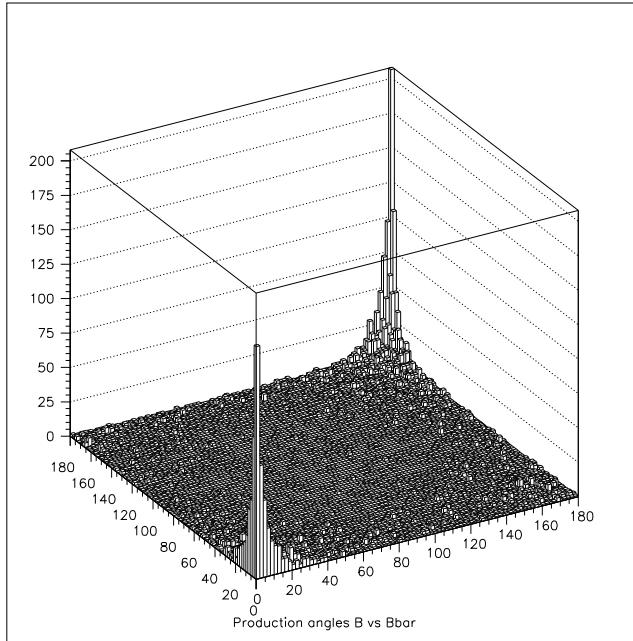
Figure 2.6: The production angle (in degrees) for the hadron containing a b quark plotted versus the production angle for a hadron containing $\overline{b}$ quark, from the Pythia Monte Carlo generator.

Thus, the forward direction at the Tevatron presents us with a number of striking advantages. First of all, there is a large cross-section for the production of correlated $b\overline{b}$ pairs. Secondly, the B hadrons that are formed have relatively large momenta, on average 30 GeV/c, and their decay products are not multiply scattered by large amounts. This allows us to make precision measurements of their spatial origins; so we can determine if they arise from B hadrons that traveled on the order of several mm prior to their decay. Furthermore the geometry is very natural for certain aspects of detector technology that significantly enhance the physics performance. For example, a Ring Imaging Cherenkov detector using a gas radiator matches the 3-70 GeV/c momentum range for B decay products. The Cherenkov photons can be detected using a relatively small area array of photomultiplier tubes or HPD's. Powerful particle identification is essential for high sensitivity b experiments. Another example is the ability to put the silicon pixel vertex detector inside the main beam vacuum. Precision detection of the B decay vertices is crucial for the trigger and in rejecting backgrounds. For these reasons, we have designed a detector with "forward coverage."

Charm production is similar to b production but has a much larger cross section. Current theoretical estimates are that charm is 1-2% of the total $p\overline{p}$ cross-section. The cross section

is even more strongly peaked in the forward direction because the average transverse momentum is of the order of only 1.5 GeV/c. The charm cross section has never been measured because experiments with good acceptance in the central region have very low efficiency for triggering and reconstructing charm. The favorable kinematics in the forward direction gives BTeV a very high efficiency for reconstructing charm.

Table 2.1 gives the Tevatron parameters which are especially relevant to BTeV design and physics reach.. We expect to start serious data taking with a luminosity of about 5×10^{31} cm^{-2}s^{-1}; our ultimate luminosity goal is 2×10^{32} cm^{-2}s^{-1}. At the higher luminosity we expect an average of 2 total interactions per crossing, 1.3 of which are not elastic or quasi-elastic.

Table 2.1: The Tevatron as a b and c source for BTeV

Luminosity (BTeV design)	2×10^{32} cm^{-2}s^{-1}
$b\bar{b}$ cross-section	100 μb
# of b's per 10^7 sec	4×10^{11}
$\frac{\sigma(b\bar{b})}{\sigma(total)}$	$\sim 0.15\%$
$c\bar{c}$ cross-section	> 500 μb
Bunch spacing	132 ns
Luminous region length	$\sigma_z = 30$ cm
Luminous region width	$\sigma_x, \sigma_y \approx 50$ μm
Interactions/crossing	$< 2.0 >$

Bibliography

[1] F. Abe *et al.*, "Measurement of the B Meson Differential Cross-Section in $p\overline{p}$ collisions at $\sqrt{s} = 1.8$ TeV," CDF/PUB/BOTTOM/PUBLIC/3759 submitted to ICHEP '96 and references theirin; F. Abe *et al.*, *Phys. Rev. Lett.* **75**, 1451 (1995).

[2] R. Abbott *et al.*, "The $b\overline{b}$ Production Cross Section and Angular Correlations in $p\overline{p}$ collisions at $\sqrt{s} = 1.8$ TeV," FERMILAB-Pub-99/144-E; S. Abachi *et al*, *Phys. Rev. Lett.* **74**, 3548 (1995).

[3] D. Fein, "Tevatron Results on b-Quark Cross Sections and Correlations," presented at Hadron Collider Physics (HCP99) (Bombay), January 1999.

[4] I. Redondo, "ZEUS Results on Charm in DIS As AN Indication of Beam Drag Effects," and E. Norrbin and T. Sjostrand, "Drag Effects in Charm Photoproduction," both in Proceedings of the workshop for **Monte Carlo Generators for HERA Physics**, 1998/99, ed. by A. T. Doyle et al., http://www.desy.de/ heramc/proceedings/wg60.

[5] M. Artuso, "Experimental Facilities for b-Quark Physics," in *B Decays* revised 2nd Edition, Ed. S. Stone, World Scientific, Singapore (1994).

[6] F. Abe *et al.*, *Phys. Rev.* D **53**, 1051 (1996).

[7] M. Mangano, P. Nason and G. Ridolfi, *Nucl. Phys.* B **373**, 295 (1992).

Chapter 3

C0 Experimental Area and Analysis Magnet

3.1   The C0 Experimental Area

The experiment will be carried out in the newly constructed C0 collision hall, shown in Figure 3.1. The hall is a 216 m^2 (9 m wide by 24 m long) enclosure centered along the C0 straight section of the Tevatron. The hall is asymmetric in the transverse direction, as best seen in the plan view, Figure 3.3, extending 3.5 m from the beam on the west side and 5.5 m from the beam on the east side. The Tevatron beams are 2.5 m above the floor slab and 4.25 m below the roof of the hall. There are flared transition enclosures both upstream and downstream. The enclosure specifications and dimensions are compatible with the detector described in this document.

Figure 3.1: C0 Collision Hall

Figure 3.2: Tevatron berm and C0 Assembly Building

To the east of the collision hall is the C0 assembly building (see Figure 3.2) a steel framed, industrial type structure containing a 150 m^2 assembly hall at the collision hall elevation. Detector components will be constructed in the assembly hall and moved into the collision hall through a 6 m × 6 m "shield door." The C0 assembly building also has space on various levels for the electrical, water and air handling systems needed for the BTeV detector. Figure 3.3 shows a layout of the assembly building and collision hall with the vertex magnet and the 4 muon toroids of the proposed BTeV detector superimposed.

Detector elements are brought into the area at a ground level loading dock and lowered to the assembly floor using the 30 ton crane which covers the loading dock and assembly hall. There is no crane in the beam enclosure; all detector elements must be designed so that they can be constructed or staged in the assembly area and moved through the shielding door into the enclosure. During operation the shielding door is sealed with a concrete door for radiation protection. The concrete door moves on rollers and is stored in the south end of the assembly area in an alcove when the door is open.

There are cable ducts from the experiment enclosure to a 150 m^2 equipment room at grade level on the north end of the assembly building. Preliminary engineering has been done on converting this equipment room into a two or three level electronics/counting room and office area for experimenters. Electric power, air handling, other utilities, as well as an elevator, lavatories, etc., although not yet installed, have been designed to accommodate BTeV needs in these areas.

3.2 The BTeV/C0 Spectrometer Magnet

The vertex magnet in the proposed BTeV spectrometer is based on an existing magnet, the SM3 magnet, which is currently part of the decommissioned Fermilab MEast Spectrometer.
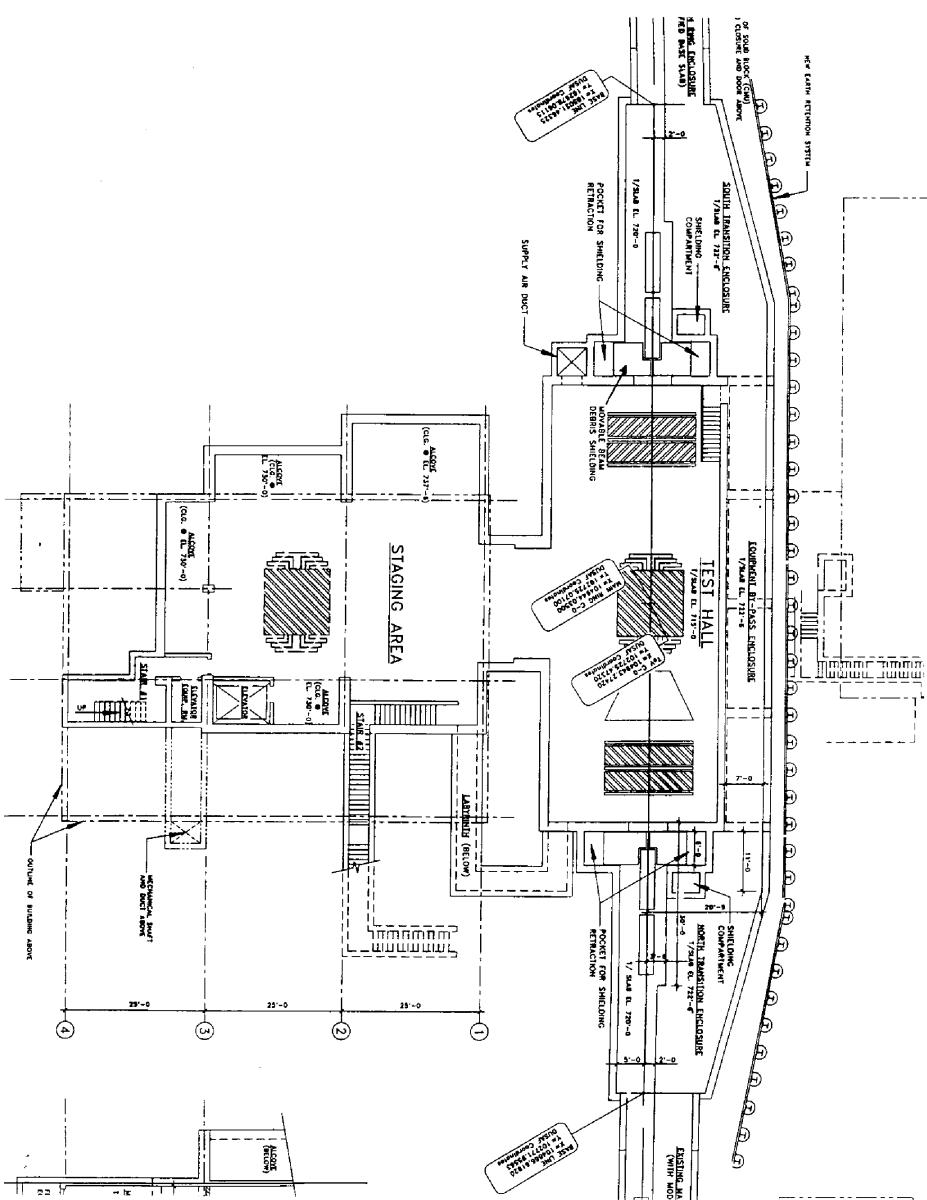
Figure 3.3: Layout of C0 Collision Area

It is shown, after modification, on the proposed layout, Figure 2.1, of the BTeV/C0 spectrometer. The SM3 magnet was assembled in 1981 from soft iron blocks that were recovered from the decommissioned Nevis Cyclotron. The coils for the magnet were built of 5 cm square aluminum conductor by the Sumitomo Corporation under the aegis of the US-Japan Agreement on High Energy Physics. The magnet operated in MEast from 1982 until 1997, at a central field of about 0.8 Tesla, serving experiments E605, E772, E789, and E866.

The SM3 magnet was assembled by welding together, in place, various blocks of the Nevis iron. It has a total weight of 500 metric tons. The construction and assembly drawings for this magnet have been located and an engineering procedure for disassembling the magnet has been developed. A test disassembly of two of the 20 large iron pieces on the magnet occurred in 1999. The test allows us to estimate the full disassembly costs with more certainty. After transportation to C0, the modified magnet will be reassembled in the C0 assembly hall and rolled into the C0 collision hall, as shown in Figure 3.3.

The central field specified for the BTeV/C0 spectrometer is much higher than the operating excitation of the existing SM3 magnet. Studies with the magnetostatic modeling programs POISSON and OPERA have led to a design for a new pole-piece for SM3. This pole-piece, indicated in Figure 3.4, yields a central field of 1.6 Tesla, and an integrated dipole field of 5.2 T-m.

The design has an integrated sextupole field component on the symmetry axis of 0.002 T/m. This is small compared to the natural sextupole moment of the Tevatron dipoles and is acceptable for insertion into the Tevatron lattice. In this design the magnet would draw 650 kW of power at 4200 amps, similar to its previous operating power levels (it was previously powered by two 500 kW Transrex power supplies). Note that the magnet will be oriented so that charged particles are deflected in the vertical plane. The properties of the magnet, with the pole faces shimmed to the BTeV requirements, are shown in Table 3.1.

Table 3.1: BTeV/C0 Vertex Dipole Properties

Property	Value	Comment
$\int B \times dl$	5.2 T-m	2.6 T-m on each side of center of IR
Central Field	1.6 Tesla	
Steel Length	3.2 m	
Overall length	5.3 m	
Magnet Vert. aperture	± 0.3 rad	
Magnet Horz. aperture	± 0.3 rad	

The magnet is centered on the interaction region in Z thus creating two forward spectrometers. In quark-antiquark production at 2 TeV, the bottom quark and antiquark are usually either both boosted in the proton beam direction, or both boosted in the antiproton beam direction. Thus, having two spectrometers doubles the acceptance of the experiment
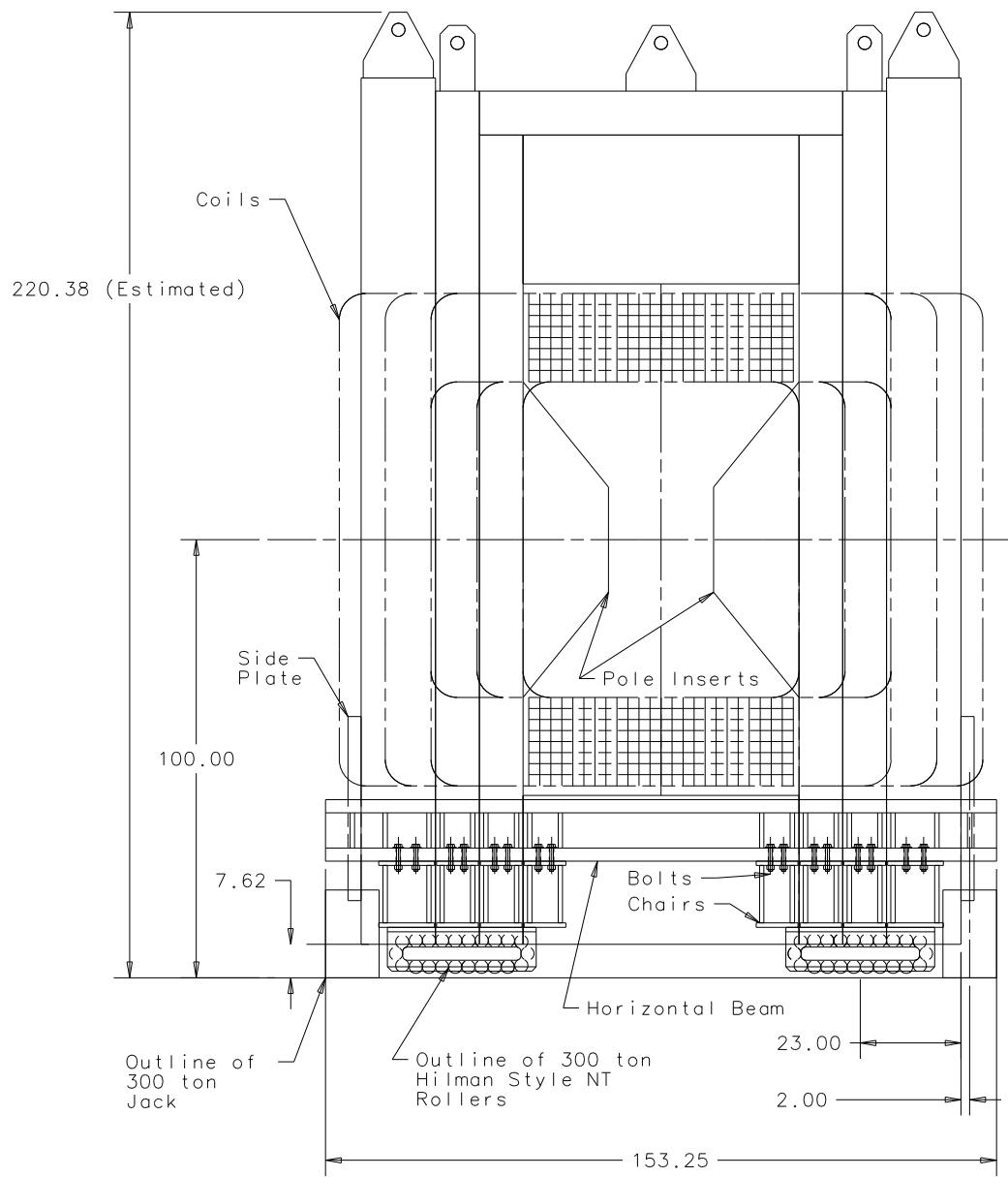
Figure 3.4: Cross section of the modified SM3 dipole with rollers and pole piece inserts. All dimensions are in inches.
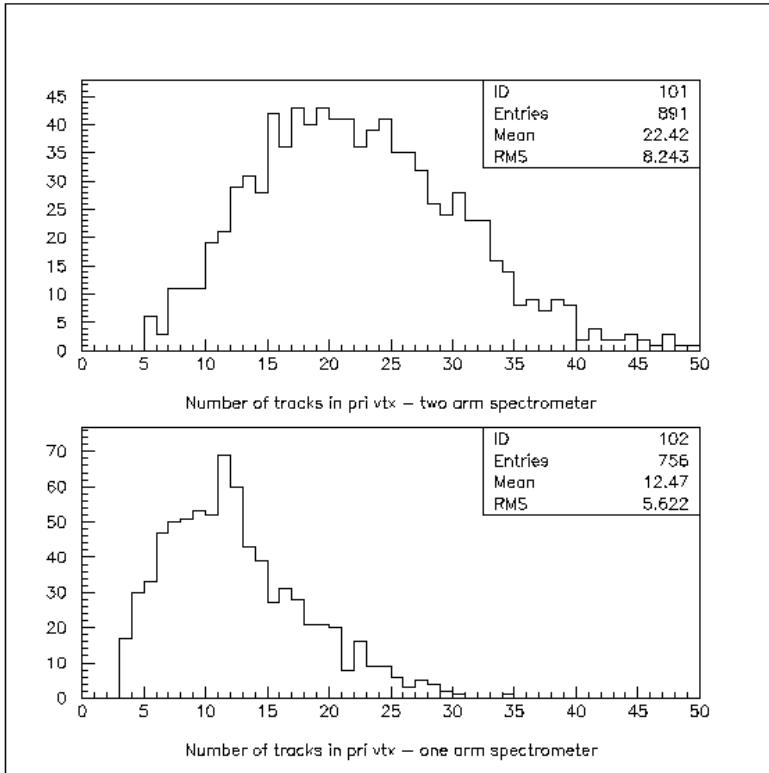
Figure 3.5: Fraction of events with N primary tracks above 1 GeV/c momentum for a) a dipole centered on the IR giving an effective two-arm acceptance and b) a single forward spectrometer arm.

for tagged decays. Having two spectrometers further improves the experiment because the extended coverage increases the number of tracks from the underlying event that can be used to determine the primary vertex. Figure 3.5 shows the number of high energy primary tracks (low multiple scattering) that contribute to locating the primary vertex for single arm coverage and for two arm coverage. If we require three such tracks to determine the primary vertex, then 2.2 times more events survive with two arms than with one, and the primary vertex resolution is enhanced.

In this central dipole geometry, there is a strong magnetic field at the vertex detector. Because of the excellent spatial resolution of the vertex detector, it is possible to get a crude measurement of the track momentum using the vertex detector alone. This measurement allows the first level trigger to properly weight tracks in vertex fits, and to impose separation cuts based on normalized miss distances.
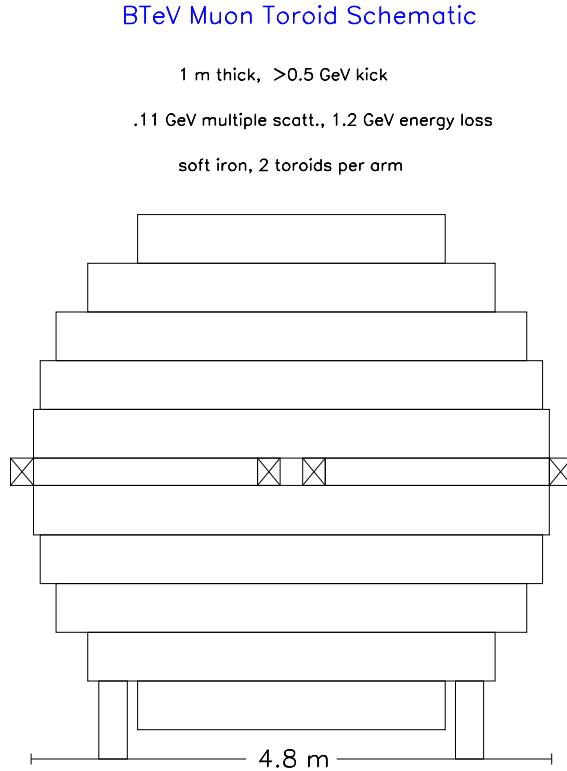
Figure 3.6: Schematic of Muon toroid.

The second spectrometer arm will also be useful during the R&D and early data-taking phases of the experiment. It is possible, given the lab's budget, that only one arm will be fully instrumented initially. The other side can be used to test new detector concepts or to test prototypes of production components under actual beam conditions until it becomes possible to instrument it fully.

3.3 The Muon Toroid Magnets

In order to absorb hadrons and simultaneously deflect muons (thus allowing momentum discrimination in the muon trigger), each spectrometer arm contains two 1 m thick magnetized iron absorber walls. Each of the 4 toroids is 4.8 m in diameter, weighs 200 tons, and has 2 excitation coils as shown in Fig. 3.6. Each toroid will be energized by 64 turns of 1000 Amp water-cooled conductor yielding a toroidal magnetic field varying from 19 kGauss at small radius to 16 kGauss at larger radius. This implies a p_T kick of >0.5 GeV/c for each toroid, or >1 GeV/c total for each muon.
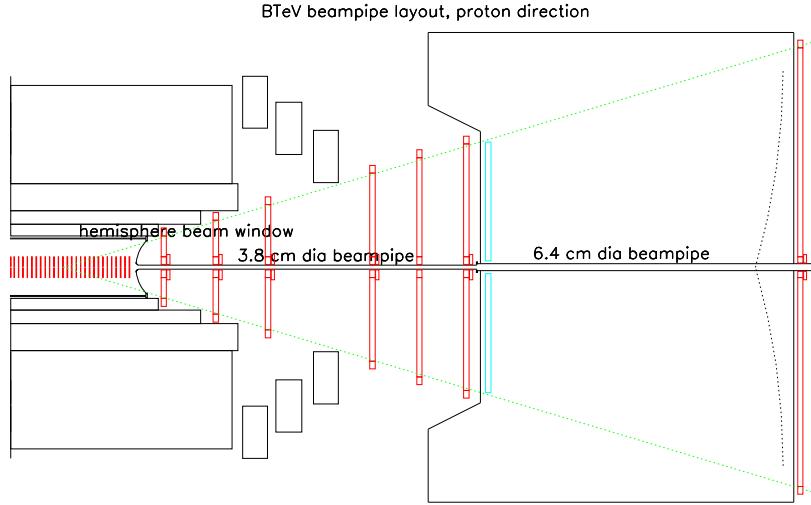
Figure 3.7: Schematic of the tracking detectors and beampipe.

3.4 The Spectrometer Beampipes

The beampipes in the regions from the end of the silicon pixel vacuum tank to the first Tevatron accelerator elements downstream of the muon detectors will be constructed of thin-walled aluminum. The beampipe will transition from the pixel vacuum box to the 3.8 cm diameter beampipe in the tracking detector region via a mushroom shaped concave dome formed from 0.030" aluminum as shown in Figure 3.7. The 3.8 cm diameter cylindrical beampipes in the tracking detector region, and the 6.4 cm diameter cylindrical beampipes in the RICH, calorimeter, and muon detector regions will also be formed from 0.030" aluminum. A special low-mass flange coupling near the upstream face of the RICH detector, coupled with a bellows section downstream of the muon detector, will allow for thermal motions and will facilitate beampipe assembly and removal.

Chapter 4

Pixel Vertex Detector

4.1 Introduction

The pixel vertex detector provides high resolution space points near the interaction, which are used both online and offline to reconstruct tracks and associate them with their parent vertices. We have chosen to use silicon pixel detectors because they provide high precision space points with very few noise hits, and they are quite radiation hard. Radiation hardness enables us to place detector elements very close to the beam (in vacuum, separated from the beam only by a thin RF shield), minimizing track extrapolation errors. The measurement of space points, with very little noise, provides superior pattern recognition, allowing us to reconstruct tracks and vertices in real time, and trigger on events containing reconstructable heavy flavor decays.

The pixel detector contains nearly thirty million rectangular pixels, each 50 μm $\times$ 400 μm. Each sensor pixel is read out by a dedicated electronics cell. The sensor pixel and the readout cell are connected by a "bump bond." The basic building block of the detector is a hybrid assembly consisting of a sensor, a number of readout chips, and a flexible printed circuit which carries I/O signals and power. The sensors are $\sim$5 cm $\times$ 1 cm, contain $\sim$25000 pixels, and are mated to $\sim$5 readout chips in a process called "flip chip assembly." These hybrid modules are supported by a moveable carbon structure that allows the pixel sensors to be positioned a safe distance away from the beamline until stable conditions have been established in the Tevatron, at which point they are moved as close to the beamline as radiation damage considerations will allow. This structure also provides cooling for the readout electronics.

4.2 Detector Specifications

The baseline vertex detector consists of a regular array of 31 "stations" of "planar" pixel detectors distributed along the interaction region (see Figure 4.1). Each station contains one plane with the narrow pixel dimension vertical, and one with the narrow dimension
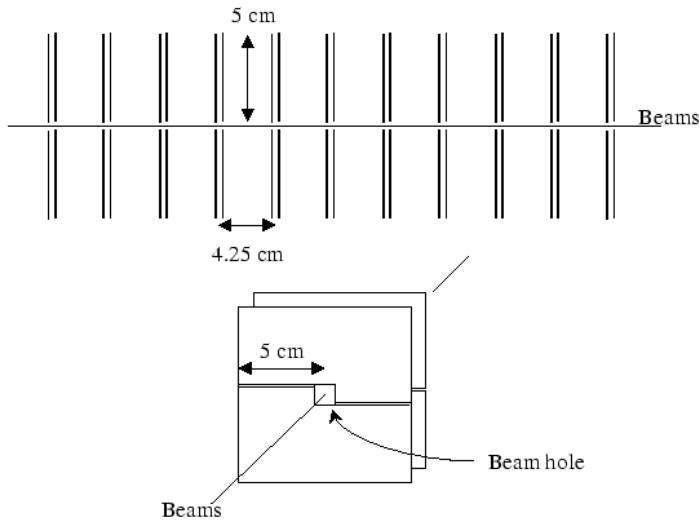
Figure 4.1: Schematic drawing of part of the pixel detector.

horizontal, and is composed of two 5 cm × 10 cm halves. The half stations are mounted above and below the beam and are arranged so that a small square hole is left for the beams to pass through. Table 4.1 summarizes the properties of the pixel detector.

4.3 Sensor, Readout-Chips, and Their Bonding

With the pixel sensors positioned to leave a 12 mm × 12 mm square beam hole, the edges of the sensors 6 mm from the beam will be exposed to a fluence of $\sim 10^{14}$ particles/cm^2 per year, given a luminosity of 2×10^{32} cm^{-2}sec^{-1}. This is similar to the fluence expected in the CMS and ATLAS pixel detectors. The issues affecting the radiation hardness of sensors are material and implant type, guard-ring design for high voltage operation, and operating temperature. The choices made by BTeV are to use n$^+$-on-n sensors with relatively low resistivity bulk material, probably with p-spray (as opposed to p-stop) n-side electrode separation, with multiple p-side guard rings, operated at about -5 °C. These choices, and the choice of pixel dimensions and signal processing requirements (e.g. the number of ADC bits required) have been made based on input from four sources. These are (1) simulation (see Appendix A), (2) BTeV experience with ATLAS sensors in beam tests at Fermilab, (3) tests on devices made in collaboration with BNL and CMS, and (4) close communication with groups world-wide working on these issues for LHC experiments. The silicon may be oxygen-diffused as prototyped by the RD42 (ROSE) Collaboration at CERN[1], depending on BTeV tests with recently received SINTEF detectors of normal and oxygen-diffused types.

Table 4.1: Pixel Vertex Detector

Property	Value
Pixel size	rectangular: 50 μm × 400 μm
Plane Dimensions	10 cm × 10 cm
Central Square Hole Dimensions (adjustable)	nominal setting: 12 mm × 12 mm
Total Planes	62
Total Stations	31
Pixel Orientations (per station)	one with narrow pixel dimension vertical & the other with narrow dimension horizontal
Separation of Stations	4.25 cm
x-Plane to y-Plane Separation (within station)	5.0 mm
Total Station Depth (incl cooling, supports)	6.5 mm
Sensor Thickness	250 μm
Readout Chip Thickness	200 μm
Total Station Radiation Length (incl RF shielding)	2%
Total Pixels	3×10^7
Total Silicon Area	$\approx$0.6 m^2
Readout	analog readout (3 bits)
Trigger	signals are used in Level 1 trigger
Rate Requirements	time between beam crossings is 132 ns.
Noise Requirement	desired: $< 10^{-6}$ per channel/crossing required: $< 10^{-5}$ per channel/crossing
Resolution	better than 9 μm
Radiation Tolerance	$> 6 \times 10^{14}$ particles/cm^2
Power per Pixel	<60 μWatt
Operating Temperature	$\sim$-5 °C

The pixel sensors will be read out by an Application Specific Integrated Circuit (ASIC) designed at Fermilab, called FPIX2. This chip will differ from pixel readout chips under development for LHC experiments in two significant ways. First, its analog front end has been optimized for the 132 ns time between beam crossings planned for the Tevatron. Second, its output bandwidth is much greater than any other chip under development; *high enough bandwidth to easily read out all hit data from every beam crossing and provide it to the Level-1 trigger hardware.* FPIX2 is being designed for fabrication in an 0.25 μm CMOS process using guard rings and enclosed-geometry transistors for radiation hardness. Recent studies by RD49[2] have shown that, given these design practices, standard deep submicron CMOS integrated circuits are at least as radiation hard as circuits implemented in military processes developed specifically for radiation hardness. In addition, the standard deep submicron CMOS is more readily available, has better yield, faster delivery times, and is much less expensive.

FPIX2 is the product of an R&D effort which started in early 1997. Two generations of prototype readout devices have firmly established the basic front-end design and readout architecture. The first generation prototype (FPIX0) contained a 12 column × 64 row array of 50 μm × 400 μm pixels. The second generation prototype (FPIX1) contained 18 columns × 160 rows of pixels of the same size. We have not yet finalized the specification for the size of FPIX2, but it may be as large as 32 columns × 256 rows (12.8 mm × 12.8 mm active area).

The silicon sensors will be connected to the readout chips using bump-bonding technology. We have had experience with indium bonding down to 30 μm pitch, and both fluxed and fluxless lead-tin solder bonding at 50 μm pitch. Both indium and fluxless solder yield acceptable results. We believe that either of these two technologies will yield a failure rate of $\sim 2 \times 10^{-4}$ or better. This belief is based on the results of our own large-scale tests with dummy circuits, and on the results obtained by other pixel R&D groups.

4.4 Multichip Assemblies and High Density Interconnections

Multichip assemblies will be made by bonding $\sim$ five readout chips (the number of chips in a module will depend on the final readout chip size) to a single sensor. Each multichip assembly will have a multilayer kapton "high density interconnect" (HDI) mounted on it. The HDI will provide connections to bias the sensor, control the readout chips, and receive output data.

One edge of each readout chip will have wire-bond pads for connection to the HDI. We have demonstrated this capability with FPIX1 chips and a prototype five-chip HDI made by Fujitsu. This prototype multilayer flexible circuit has 20 μm lines, and 20 μm line-to-line separation. Vias from an outer layer to an inner layer in the flex circuit have a pad size of 108 μm (diameter) and a hole size of 25 μm. Vias between the inner layers have a 350 μm pad size and a 25 μm hole size. Electrical tests of the five-chip assemblies have demonstrated that the prototype five-chip module performs at least as well as our previous single chip assemblies.

In our baseline design, the HDI will carry signals between the FPIX2's and the control and readout circuitry located at one end of the HDI. This circuitry will include a radiation hard data serializer, a VCSEL optical driver, a PIN diode optical receiver, and a control, monitoring, and timing ASIC. The VCSEL and the optical receiver are commercial radiation hard parts. We plan to use a radiation hard data serializer ASIC (CHFET) developed for CMS, which we have tested. The control, monitoring, and timing ASIC will be developed at Fermilab. We are also pursuing the possibility that the FPIX2 chips can drive data directly (using high speed LVDS over copper) to commercial FPGA's located far enough from the beamline so that they need not be radiation hard. In this scenario, the control and monitoring functions would also be performed by commercial devices located outside the high radiation area.
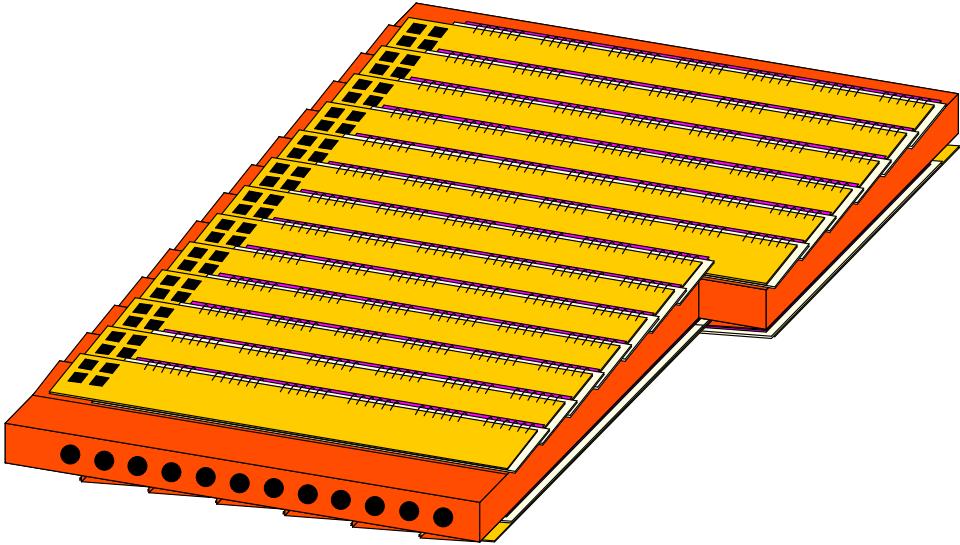
Figure 4.2: Pixels on the half plane.

Figure 4.2 shows a sketch of a half-plane assembly.

4.5   Cooling Support, Shielding, and Housing

The pixel half-planes will be "L-shaped," with one frame above and one below the beam. Each frame will be movable such that the alignment with respect to the beam is adjustable. In that way, the size of the beam hole will also be adjustable, separately both horizontally and vertically. This will allow us to optimize the balance between radiation damage and physics reach. Smaller hole size gives better vertex resolution, acceptance, tagging efficiency, and thus physics reach. The experiment performance can be optimized as experience is gained, both in operating conditions and physics analysis issues.

The mechanical supports will be carbon composite structures which will include integrated cooling channels (tubes). These structures will be manufactured by Energy Science Laboratories, Inc. (ESLI) by sandwiching nonporous carbon tubes between two sheets of carbon "flocking," which consists of many individual carbon fibers and looks like velvet material. Additional carbon will then be vapor deposited into the structure, making a rigid composite. The "fuzzy carbon" surfaces of this structure will be machined to provide a "shingled" surface on which the multichip assemblies will be mounted. This construction allows sensors on one side of the cooling support to overlap, and therefore allows two pixel measurement planes to be mounted on a single cooling support. Heat is transferred through the thickness of the readout chips to the carbon fibers of the fuzzy surface. These carbon fibers efficiently transfer the heat to the coolant flowing in the cooling tubes. ESLI has delivered a first prototype for thermal and mechanical testing (see Figure 4.3).

Figure 4.3: Sample fuzzy carbon composites; note the shingling in the picture on the right.

It is likely that the pixel detector and readout hybrid assemblies will need to be electrically shielded from the circulating beams. The HERA-B vertex detector currently operates successfully with a 150 μm thick aluminum RF shield[3]. We have built, and successfully operated, a test facility to simulate the Tevatron beam (using an RF generator and a wire inside a stainless steel pipe) to allow us to understand the detailed needs of the BTeV pixel system (see Appendix A).
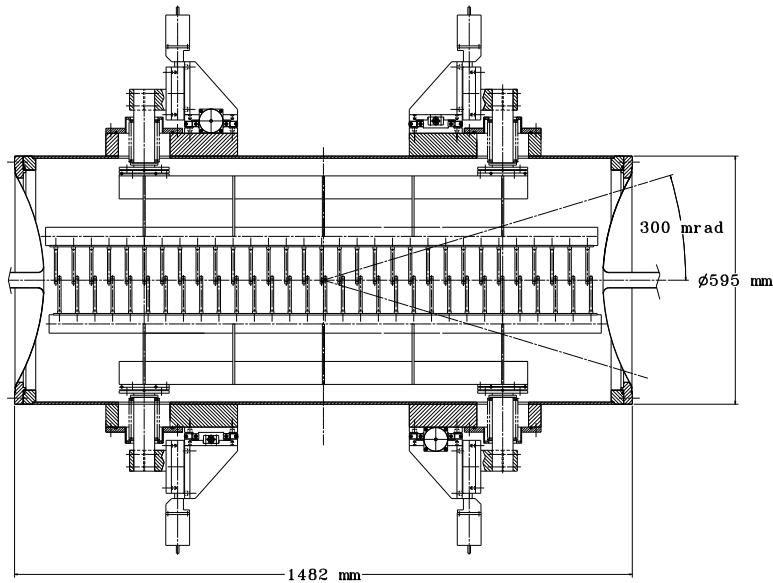

Figure 4.4: Side view of the vacuum vessel and support structure for the pixel detector.

Figure 4.4 shows a conceptual design for the aluminum vacuum vessel and carbon support structure for the pixel detector. The vessel is cylindrical with a length of $\sim$ 150 cm and a

66

diameter of 59.5 cm. Particles within the 300 mrad acceptance of the spectrometer traverse only the pixel stations and the 0.75 mm thick exit window. The carbon support and cooling planes will be held in position by a structural cooling manifold, whose position will be controlled by motors located just outside the vacuum vessel. The design and orientation of these motors will allow for operation in the ~ 1 T field of the spectrometer magnet at the motor locations.

The expected alignment precision of the pixel assembly will be ~ 1 μm, a level which has been achieved by HERA-B[4]. Final alignment, of course, will be done with particle tracks, with a rapid-turnaround computing system dedicated to obtaining and checking the alignment parameters for the trigger at the start of each fill. Note, however, that the interaction vertex is determined event-by-event in the same coordinate system as the secondary vertices. Thus, the exact position of the pixel array with respect to the beam need not be reproduced fill to fill.

4.6   Total Material

$Item$	$Thickness(mm)$	$X_0(mm)$	$Coverage$	$X/X_0(\%)$
$Sensor$	0.25	93.6	1.20	0.32
$Readout\ chip$	0.20	93.6	1.15	0.25
$Bumps\ and\ wire\ bond$	0.02	10.0	0.02	0.004
HDI	0.224	284	1.00	0.08
$Components\ on\ HDI$				0.04
$Glue$				0.02
$Coolant(water/alcohol)$	0.35	400	0.10	0.01
$Substrate(C-C)$	0.5	427	0.50	0.06
$Shielding(Al)$	0.1	89.0	1.00	0.11
$Total$				0.89

Table 4.2: Material budget of a BTeV pixel plane. The column labelled "coverage" shows the factor applied to account for overlaps of the sensors and readout chips, and for geometric coverage (e.g. area covered by bump bonds/total area). The numbers given for components on the HDI and for glue are derived from the ATLAS Pixel Detector TDR [5].

Table 4.2 summarizes the material budget for one pixel plane. The numbers for the support and cooling structure for a full station, and for the coolant, have been divided by two (coverage column), since two planes are held on one support. To account for the non-sensitive area in each module (required for guard rings and scribe edge for the sensor, and for the periphery and wire bond pads of the readout chip), a 20% overlap of sensors and a 15% overlap of readout chips are included.

4.7 Test Beam Results

During the 1999 Fermilab fixed target run, tests were performed using both FPIX0 and FPIX1 readout chips bonded to ATLAS prototype pixel sensors (see Appendix A). Data was taken with as many as four pixel devices in a telescope configuration. These four pixel planes, and smaller subsets of them, were placed between upstream and downstream silicon-microstrip telescope elements. Thus, the pixel devices could be separately studied as they might be used in an experiment, as well as compared to beam-track projections from the microstrip system. Resolution was studied as a function of the angle between the beam and the pixel plane. In a short test, a diamond target was placed upstream of the four plane pixel telescope, and multi-particle interactions were recorded and analyzed (see Figure 4.5). In addition, tests were done with a pixel plane in a magnetic field while beam particles traversed the pixel plane and upstream telescopes. Each pixel configuration was tested with a variety of applied sensor biases and readout thresholds to study the sensitivity of the results to these parameters.
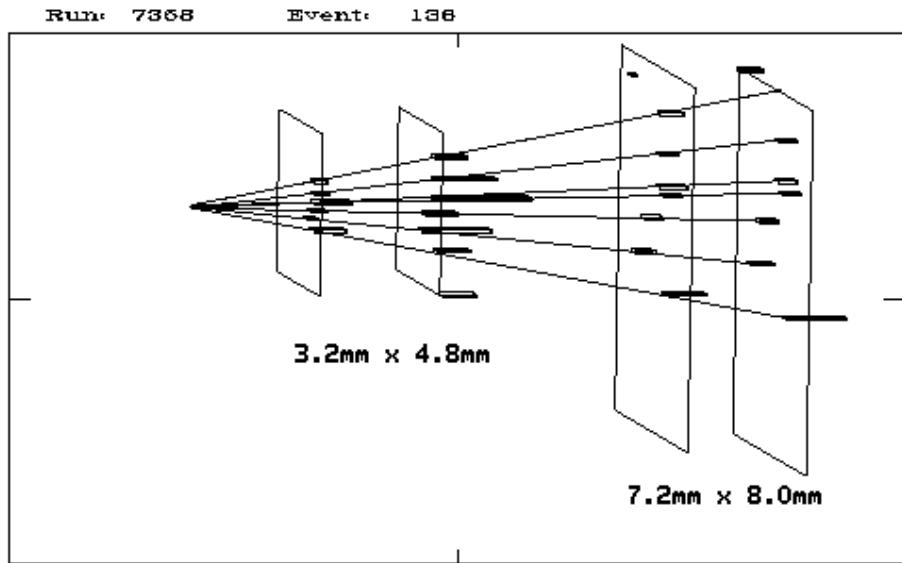
Figure 4.5: Multiparticle interaction observed in Fermilab beam test. This density of tracks is much higher than expected in the BTeV pixel detector.

Figure 4.6 shows the resolution (in the coordinate measured by the 50 μm pixel pitch) as a function of the incident beam angle for an FPIX0-instrumented detector. The solid line in the plot is the (Gaussian) resolution function used for the Monte Carlo studies presented in this proposal. This line is a piecewise linear fit to the resolution obtained in BTeV pixel simulations. FPIX0 provided an analog output which was digitized by an off-chip 8-bit FADC. The plot shows both the resolution obtained using this 8-bit information directly,
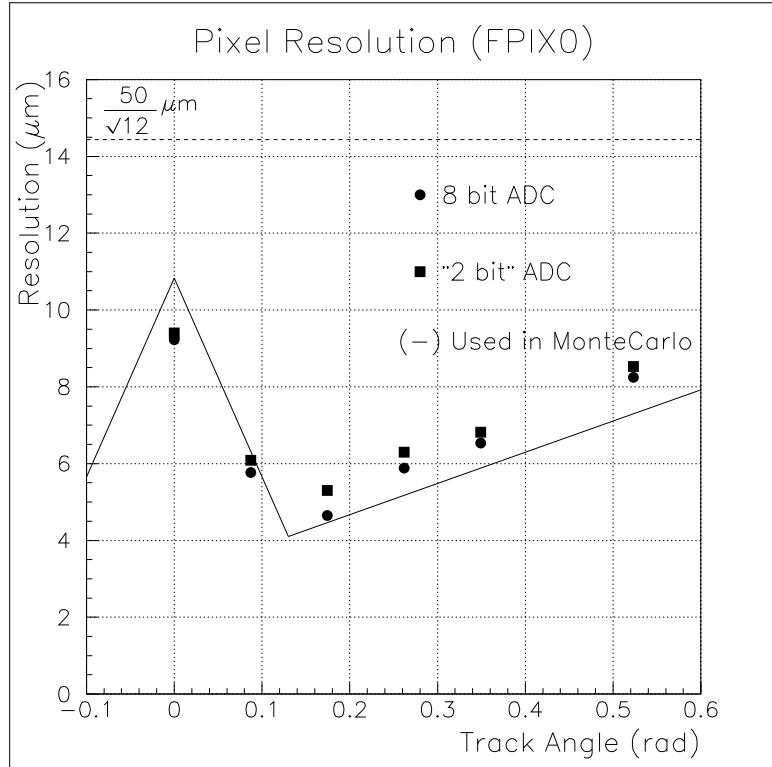
Figure 4.6: Resolution as a function of the angle of the incident beam.

and also the resolution obtained by degrading the pulse height to 2-bits of information. This result confirms our simulation result that very good resolution can be obtained using charge sharing even with very coarse digitization. Based on these results it has been decided that FPIX2 will have a 3-bit FADC in each pixel cell. This will provide excellent spatial resolution. In addition, since the ADC will be implemented using eight independent comparators, some of the thresholds may be used to indicate the presence of delta rays and photon conversions.

We have also used test beam data to parameterize the non-Gaussian tails (due to low energy delta rays) in the pixel detector resolution function. This parameterization has been used in Monte Carlo simulations of the vertex trigger.

These test beam results show several important things:

- ATLAS prototype sensors coupled to FPIX chips meet BTeV resolution requirements

- The FPIX-type front-end performs as needed

- The BTeV multiplicity is easily handled by 50 μm × 400 μm pixels

- Alignment and support stability are important to meet the resolution goals

4.8 Summary

Over the last three years, BTeV collaborators have been working vigorously to establish a pixel detector capability for Fermilab, tuned to the unique features of the Tevatron crossing time and BTeV trigger needs. Since the earliest concerns were related to sensors, readout, and bump bonding, the primary focus has been on those issues. Progress has been gratifyingly rapid, both at Fermilab and for the LHC. This progress is evident in the success of the test beam effort at Fermilab, the results of which validate the ideas used for this BTeV proposal. The proposal to use a "shingled" fuzzy carbon support and cooling structure appeared recently, and preliminary cooling capacity tests are underway on a prototype. This is the necessary first step needed to validate the material estimate presented here. Yields of all pixel-related components have been much higher than anticipated, giving hope that a focus on simplicity within our aggressive technical approach will succeed. This approach, combined with the early implementation of a significant subsample ($\sim$10%) of final-design detectors, should allow reliable planning and achievable goals for the experiment.

Bibliography

[1] F. Lemeilleur, *et al.*, "3rd RD48 Status Report", CERN/LHC 2000-009 (31 December 1999). Available at http://rd48.web.cern.ch/RD48/status-reports/RD48-3rd-status-report.doc (requires MS Word)

[2] P. Jarron, *et al.*, "3rd RD49 Status Report Study of the Radiation Tolerance of ICs for LHC"; CERN/LHCC 2000-03 (13 January 2000). Available at http://www.cern.ch/RD49/RD49News/RD49StatusReport3.pdf (requires Acrobat Reader 4)

[3] C. Bauer *et al.*,"First Experience and Results from the HERA-B Vertex Detector System"; Nucl. Instrum. Meth. A 418, (1998) 65.

[4] Iris Abt, private communication.

[5] "ATLAS Pixel Detector Technical Design Report", CERN/LHCC 98-13, available at http://atlasinfo.cern.ch/Atlas/GROUPS/INNER_DETECTOR/PIXELS/tdr.html

Chapter 5

Forward Tracking System

5.1   Introduction

The major functions of the forward charged particle tracking system are to provide high precision momentum measurements for tracks found in the pixel system, to reconstruct and measure all parameters for tracks which do not pass through the vertex detector (such as K_s and Λ^0 daughter tracks), and to project tracks into the RICH counters, EM calorimeters, and Muon detectors. Measurements from the forward tracking system are also used online in the Level 2 trigger, as explained in Chapter 8.

5.2   General Description

The baseline forward tracking system consists of 14 stations, 7 in each arm, placed transversely to the beam at various distances from the interaction point. Three stations are placed in the fringe-field region of the dipole magnet, three stations in the field-free region just upstream of the RICH, and one station just downstream of the RICH. The entire system extends over a distance of $\sim \pm 7$ m and provides θ-angle coverage from $\sim \pm 10$ mrad up to ± 300 mrad.

The design of the forward tracking system has been driven by the high density of tracks produced in the forward direction, especially with multiple interactions per crossing. Two different types of detectors are used. Most of the solid angle is instrumented using straw tube drift chambers. Straws have been chosen because they can be used to make large chambers with small cell size, and because they can be built to surround the beam pipe without requiring a heavy frame near the beam. The track density very close to the beam requires detectors with even higher granularity; we have chosen to instrument the central section of each station with silicon microstrip detectors.

Tables 5.1 and 5.2 list all the geometric parameters and the main characteristics of the forward tracker.

This forward tracking system configuration has sufficient segmentation to handle the

Table 5.1: Properties of the baseline forward straw tracker

Property	Value
Straw size	4 mm diameter
Central hole	24 cm × 24 cm
Total Stations	14 (7 per arm)
Z positions (cm)	96, 146, 196, 296, 341, 386, 706
Half size (cm)	30, 45, 60, 90, 105, 118, 210
Views per station	3 (X,U,V)
Layers per view	3
Total number of straws	66,780
Total station thickness	0.6% X_0
Total channels	118,440
Readout	ASD + timing chip (6 bits), sparsified

Table 5.2: Properties of the baseline forward silicon tracker

Property	Value
Si-sensors	6" wafers, p-on-n type
Pitch	100 μm
Thickness	200 μm
Sensor configuration	4 ladders of 2 sensors + 2 single sensors
Coverage	24 cm × 24 cm
Central hole	5.4 cm × 5.4 cm (7 cm × 7 cm in last station)
Total stations	12 (6 per arm)
Z positions (cm)	100, 150, 200, 300, 390, 710
Views per station	3 (X, U, V)
Channels per view	3,000
Total channels	108,000
Readout	sparsified binary

high hit multiplicities that are expected when $b\bar{b}$ events are produced in the forward region. Fig. 5.1 shows occupancies in the straw tracker predicted by BTeVGeant for the case in which a $b\bar{b}$ event is produced at the design luminosity of $2{\times}10^{32}$ cm^{-2} s^{-1}. It is worth noting that these occupancies are almost a factor of two higher than those expected on average at this luminosity. The maximum occupancy in the silicon strip detectors, which have 40 times finer pitch than the straw chambers, is everywhere less than ~4% for the conditions of Fig. 5.1.
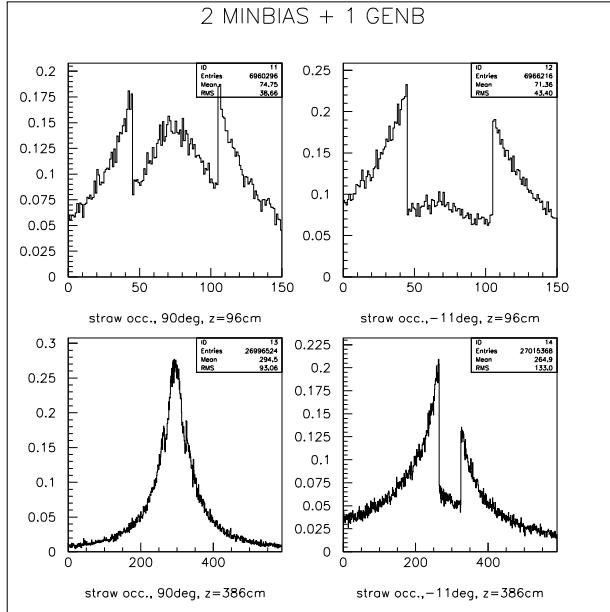
Figure 5.1: Occupancies in the first station of straws, and the station just upstream of the RICH counter, when a $b\bar{b}$ event is produced at the design luminosity of 2×10^{32} cm^{-2} s^{-1}. The two histograms on the left are for X-view straws, while those on the right are for U-view straws. The V-views have identical occupancies (mirrored about zero) to the U-views.

5.3 Forward Straw Tracker

5.3.1 Detector Description and Layout

The forward straw tube tracker consists of stations that provide 3 coordinate measurements, X, U and V, where the two stereo views, U and V, are at $\pm 11.3^o$ around the Y bend coordinate. With three layers per view, this configuration provides excellent resolution in the bend plane while maintaining a robust ability to reject ghost combinations of hits. It has sufficient redundancy to achieve a high detection efficiency and to resolve the left/right ambiguity a very large fraction of the time. The construction is modular, following a technique developed for the SDC straw tracker[1].

All the sense wires for the straw cells that do not terminate at the central hole are divided electrically using a small glass capillary bead following the technique used for the ATLAS TRT [2]. This cuts the occupancy rates in half. The sense wires in straws that span more than 80 cm have additional supports, which are realized following the helical design developed for the ATLAS TRT.

5.3.2 Front End Electronics and Drift Time Measurement

The straw tube chambers will be instrumented using electronics developed by the University of Pennsylvania [3], initially for the SDC straw chambers, and more recently for the ATLAS TRT. These radiation hard integrated circuits include high gain preamplifiers, pole-zero networks for pulse shaping and ion-tail cancellation, and leading edge discriminators.

The drift time will be measured using digital TDC's. The information from the straw tracker, like all information from every subsystem in the BTeV spectrometer, must be digitized and read out for every crossing. This means that a new TDC must be designed for BTeV. The small diameter of the straws makes the specifications of this TDC easy to achieve. A six-bit single-hit TDC, with 1.5 ns wide bins covering 96 ns, is sufficient to provide a drift distance measurement precision better than 100 μm.

5.3.3 Technical issues

We are developing a prototype straw tube which places an aluminum conduction layer between two Kapton films, the inner one next to the gas volume being a carbon loaded, low resistivity film. The idea is that the Kapton forms a protective barrier, similar to the graphite layers deposited on the inner surface of the ATLAS TRT straws. Without this protective barrier, there is a danger that the aluminum layer may be etched away, limiting the lifetime of the straw. We measured the surface resistivity of the aluminum coated, carbon loaded Kapton film of our prototype to be 6.5 ± 1.0 Ω/square, which is comparable to the specified value for the TRT straw tube. The details of the prototype straw material are listed in Table 5.3. The 1 mil thickness of each film is chosen as a compromise between the 0.5 mil thick films used for SDC straw tubes (which had very little mechanical rigidity) and the thicker and reinforced TRT straws. Our final design will likely have the straws supporting some of their own weight. If this turns out to not to be the case, we may use instead 0.5 mil Kapton films which have been shown to work and would provide a reduction of material in the detector volume.

Description	BTeV Straw Prototype
Kapton film	Inner: Polyimide type XC 25 ± 2.5 μm thickness Outer: Polyimide type 100 VN 25 ± 2.5 μm thickness
Density	1.42 g/cm^3
Aluminum layer	$(0.2 \pm 0.08)\mu m$ thickness
Resistivity of inner Kapton layer	6.5 ± 1.0 Ω/square

Table 5.3: Summary of material specifications for the BTeV prototype straw tubes

5.4 Forward Silicon Tracker

5.4.1 Detector Description and Layout

Our design consists of stations with three planes of 200 μm thick single-sided silicon microstrip detectors with 100 μm pitch. On each plane, ten silicon detector wafers are mounted on low mass carbon fiber support and are arranged, wherever possible, in ladders of two daisy-chained Si-sensors to minimize the number of readout channels. The resulting configuration is depicted in Fig. 5.2.
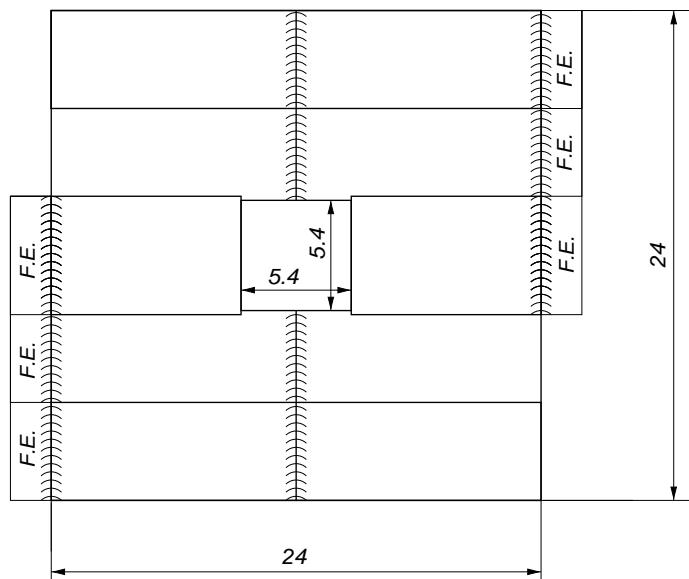
Figure 5.2: Sketch of a silicon detector plane. It consists of 4 ladders of two daisy-chained Si-sensors plus 2 single sensors. One strip overlap between contiguous detector elements ensures good efficiency over the entire plane. Dimensions are in centimeters and *F.E.* stands for Front-End read out electronics.

Three views are provided by rotating the planes by the appropriate angles: X, U and V, where the two stereo views, U and V, are at $\pm$ 11.3^o around the Y bend coordinate. Each plane consists of about 3,000 channels; the entire system of 12 stations, 6 in each arm, has about 108,000 channels in total.

The Si-sensors, having a length up to 12 cm, are of the standard p-on-n type, with multiple n-side guard rings to allow high voltage operation, and are produced using 6" wafer technology.

The front-end electronics is distributed along the two opposite sides of each plane where it is cooled by a fluid circulating in a duct embedded in the support structure all around the periphery of the plane. This is also enough to keep the sensors at low temperature, provided that the sensors and the support are in close thermal contact. In our structure, we ensure good thermal conductivity all across the planes by employing long sensors.

The preamplifiers are AC coupled to the strips by means of capacitors directly integrated on the sensors. Each channel is read out in binary mode providing a $\sigma = 100 \ \mu\mathrm{m}/\sqrt{12} = 29$ μm resolution, adequate for our physics goals.

We do not foresee any major problems in building these detectors since we can profit from the enormous experience accumulated in CDF, as well as in other experiments, and that coming from the ongoing R&D programs for LHC experiments. Nevertheless, we anticipate a possible minor concern that requires a proper backup solution. If the production of 200 μm thick 6" wafers would be problematic, we will use ladders of shorter sensors wherever necessary.

5.4.2 Radiation Issues

It is well known that the exposure of silicon detectors to high radiation doses causes damage that limits their useful lifetime. Thanks to the enormous progress accomplished during the last few years, we can now build detectors that can be operated after exposure to fluences in excess of 10^{14} particles/cm^2 [4].

In BTeV, we expect a radiation level at the silicon detectors that decreases rapidly with increasing distance from the beam. Important radiation damage effects, if any, will be confined to a small region around the central hole of the stations.

The highest levels of radiation occur at the station closest to the interaction region in the two symmetric arms of the apparatus. As shown in Fig. 5.3, the maximum value of the fluence is expected to be $\sim 1.6 \times 10^{13}$ particles/cm^2/year, given a luminosity of 2×10^{32} cm^{-2} s^{-1}. This is slightly less than the dose expected for Layer 0 of the CDF silicon tracker at the same luminosity [5]. With a proper choice of detectors, and by keeping them at low temperature, such as 0^oC, we will operate the detectors with a safety margin superior to that of CDF and those of LHC experiments. In the worst case scenario, we can expect serious radiation damage effects only on a minor portion of our detectors close to the beam after several years of operation.

We are starting an R&D program on silicon sensors to investigate other possibilities to further extend the useful lifetime of the forward silicon tracker. In particular, oxygenated sensors look very promising.

5.4.3 Readout Electronics

Even given the low occupancy expected in the Forward Silicon Tracker, the output bandwidth required to read out all hit information from every crossing is higher than is provided by any SSD chip, either already fabricated or being developed for another experiment. For this reason we have decided to develop a new readout chip with very high readout bandwidth. We will also take the opportunity to design a continuous-time-filter preamplifier capable of exploiting all the advantages offered by the relatively long bunch-crossing period of the Tevatron collider (T = 132 ns). We are considering a new preamplifier that is derived from the BaBar silicon strip front-end. We can anticipate an ENC $\sim 1000 \ e^-$ for semi-Gaussian
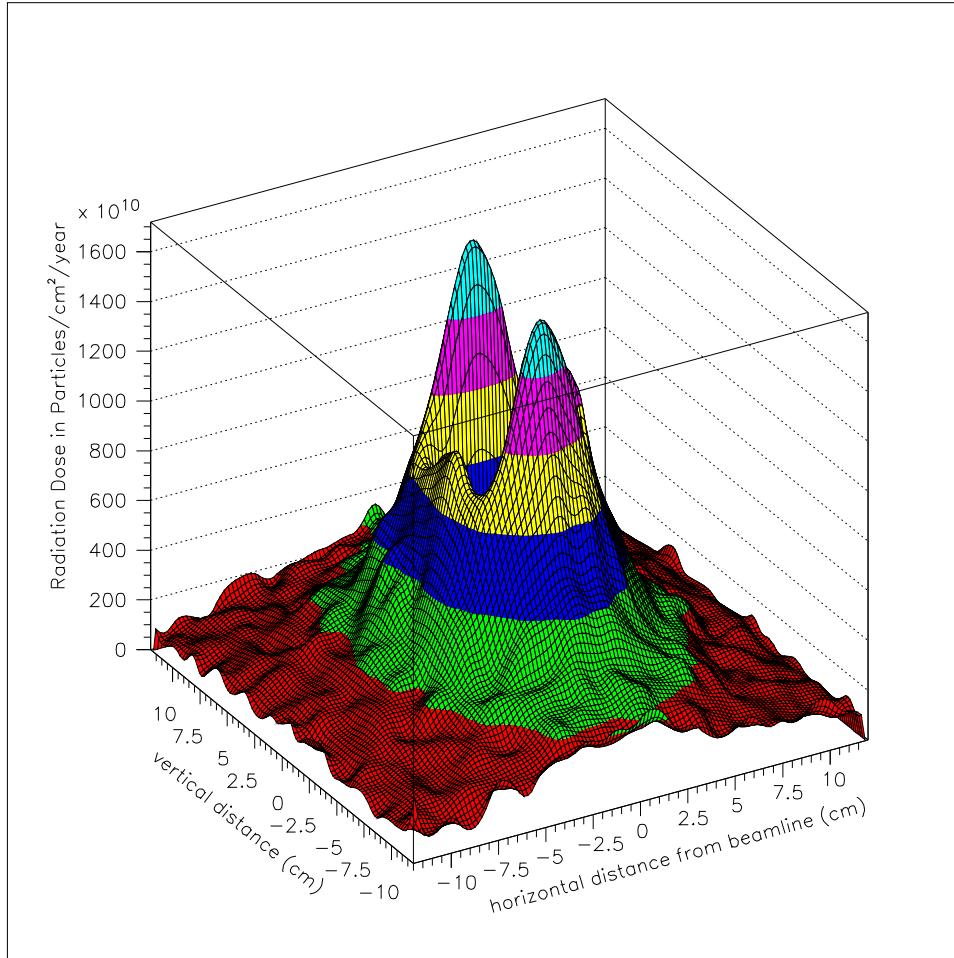
Figure 5.3: Radiation dose as a function of position in Forward Silicon Tracker Station #
1. The horizontal magnetic field concentrates more particles above and below the square
central beam hole than on either side.

shaping with 100 ns peaking time and a capacitive load at the input of $\sim$30 pF, as expected
for our longest strips. This noise performance represents in our view "the state of the art"
for silicon strip preamplifiers.

The binary readout we are presently considering is a simplified version of the readout
scheme implemented in the FPIX2 pixel readout chip. It is very fast and employs a flexible
scheme for zero-suppression and readout, that can be easily adapted to strips. The SSD
readout chips will be designed to interface to the same electronics we will employ to read
out pixel chips.

We have initiated an R&D program to design this chip, which will be implemented using
0.25 μm CMOS, following the radiation tolerant design rules developed for the FPIX2 design.

5.5 Forward Tracker Performance

The system just described ensures excellent tracking performance over the full acceptance of the forward spectrometer. Figures 5.4 and 5.5 show the expected average fractional momentum resolution for b decay products as a function of track momentum and of the track production angle with respect to the beam axis. For these histograms, an effective position resolution of $\sigma_{X,U,V} = 150$ μm was assumed for each view of the straws and a resolution of $\sigma_{X,U,V} = 29$ μm assumed for the silicon strip detectors.
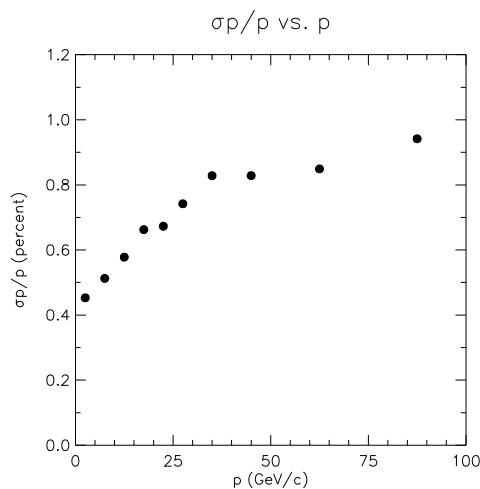
Figure 5.4: Momentum resolution as a function of track momentum for b decay products.
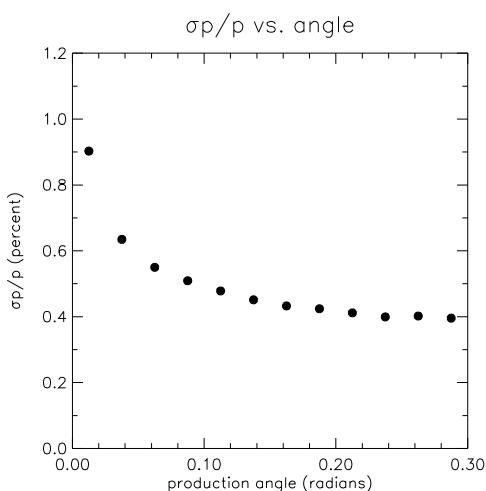
Figure 5.5: Momentum resolution as a function of polar production angle for b decay products.

Bibliography

[1] Y. Arai, *et al.*, "A modular straw drift tube tracking system for the Solenoidal Detector Collaboration experiment, Part I: Design"; Nucl. Instrum. Meth. A 381 (1996) 355.

[2] ATLAS Inner Detector Technical Design Report, CERN/LHCC/97-16,17.

[3] F.M. Newcomer, R. Van Berg, J. Van der Spiegel and H.H. Williams, Nucl. Instrum. Meth. A 283 (1989) 806.

[4] G. Tonelli, *et al.*, "The R&D program for silicon detectors in CMS"; Nucl. Instrum. Meth. A 435 (1999) 109.

[5] M.A. Frautschi, "Radiation Damage Issues for the SVX II Detector"; CDF/DOC/SEC_VTX/PUBLIC/2368.

Chapter 6

Charged Particle Identification System

Excellent charged hadron particle identification is a critical component of a heavy quark experiment. Even for a spectrometer with the excellent mass resolution of BTeV, there are kinematic regions where signals from one final state will overlap those of another final state. For example, $B_s \rightarrow D_s K^-$ signal must be distinguished from $B_s \rightarrow D_s \pi^-$ background in order to measure the CKM phase γ. These ambiguities can be eliminated almost entirely by an effective particle identifier. In addition, many physics investigations involving neutral B-mesons require "tagging" of the flavor of the signal particle by examining the properties of the "away-side." Our studies show that kaon tagging is a very effective means of doing this. "Same-side" kaon tagging is also very effective for B_s mesons.

6.1   Requirements

In the design of any particle identification system, the dominant consideration is the momentum range over which efficient separation of the various charged hadron types – π, K, and p – must be provided. In BTeV, the physics goal which sets the upper end of the momentum requirement is the desire to cleanly separate $B_d^o \rightarrow \pi^+\pi^-$ from $B_d^o \rightarrow K^+\pi^-$ and $B_s^o \rightarrow K^+K^-$. These two-body decays produce reasonably high momentum pions and kaons. Fig. 6.1 shows the momentum distribution of pions from the decay $B_d^o \rightarrow \pi^+\pi^-$ for the case where the two particles are within the spectrometer's acceptance. The low momentum requirement is defined by having high efficiency for "tagging" kaons from generic B decays. Since these kaons come mainly from daughter D-mesons in multibody final state B-decays, they typically have much lower momentum than the particles in two body decays. Fig. 6.2 shows the momentum distribution of tagging kaons for the case where the signal particles are within the geometric acceptance of the spectrometer. About 1/5 of the tagging kaons never exit the end of the spectrometer dipole. Almost all kaons exiting the dipole have momenta above 3 GeV. Based on these plots, we have set the momentum range requirement for the particle identification system to be

$$3 \,\text{GeV}/\text{c} \quad < \quad P_{particle\ id} \quad < \, \sim 70 \,\text{GeV}/\text{c} \tag{6.1}$$

Finally, kaons and pions from directly produced charm decays have momenta which are not very different from the kaons from B-decays. The range set by the B-physics requirements is a reasonable, if not optimal, choice also for charm physics.
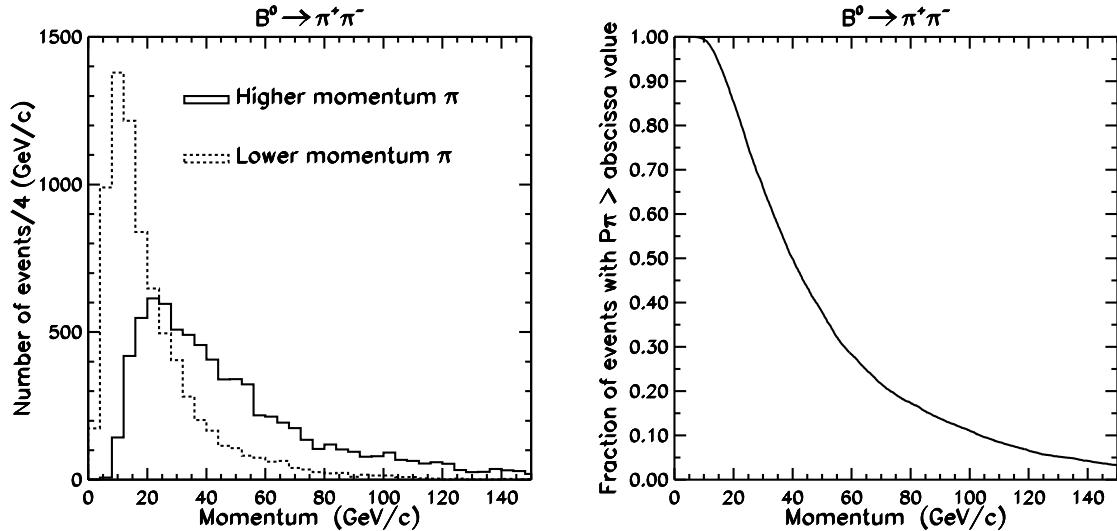

Figure 6.1: The momentum distribution of pions in $B_d \to \pi^+\pi^-$ decays. The left plot shows distributions for the lower (dashed line) and higher (solid line) momentum pion in this decay. The right plot presents the latter distribution in integral form, which gives loss of efficiency as a function of the high momentum cut-off of the particle ID device.

6.2 RICH radiators

Because of the large momentum range and limited longitudinal space available for a particle identification system in the C0 enclosure, there is really only one choice of detector technology – a gaseous ring-imaging Cherenkov counter. Pions and kaons can be separated in this momentum region with a single gas radiator. We choose C_4F_{10} which has an index of refraction of 1.00138 in the visible range. The momentum dependence of the Cherenkov angle for pions, kaons and protons in this gas is shown in Fig. 6.3. Many other experiments use this gas, including DELPHI (endcap) [1], HERA-B [2] and HERMES [3]. It was also chosen for one of the LHC-b RICH detectors [4].

Note that below about 9 GeV, no gas can provide K/p separation and that, since kaons are below threshold, the RICH operates in a threshold mode for K/π separation (except that it has much better noise discrimination than a normal threshold counter because it still measures a Cherenkov ring for pions). In order to improve identification of low momentum particles, one interesting possibility is to insert a thin ($\sim$ 4 cm) piece of aerogel at the entrance to the gas RICH as proposed by LHC-b [5] and already implemented by HERMES [3]. For example, aerogel with a refractive index of $n = 1.03$ would lower the π, K, p
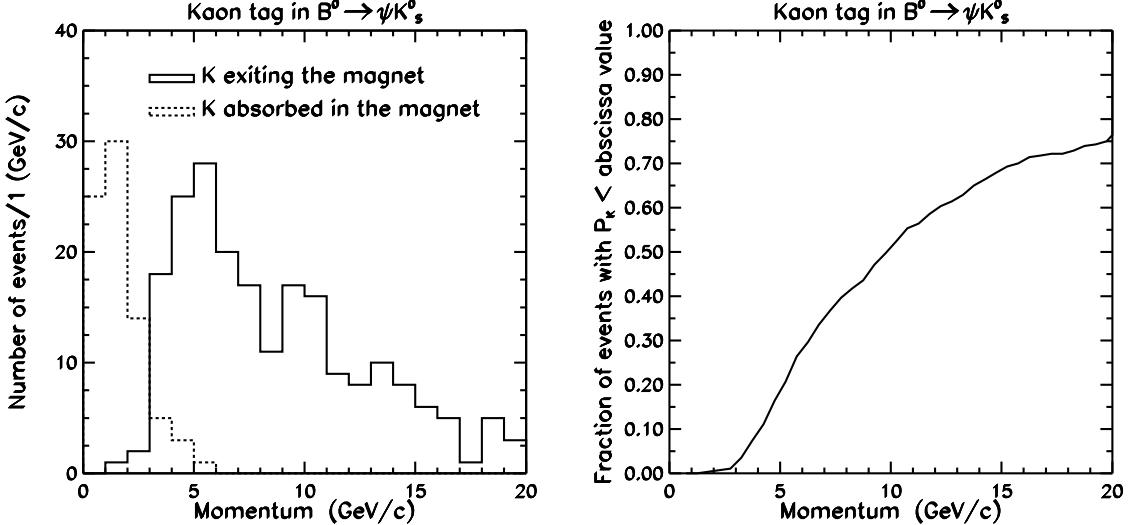
Figure 6.2: The momentum distribution of "tagging" kaons for the case where the signal particles (ψK^0_S) are within the geometric acceptance of the spectrometer. The left plot shows distributions for kaons absorbed in (dashed line) and exiting from (solid line) the magnet. The right plot presents the latter distribution in integral form, which gives loss of efficiency as a function of the low momentum cut-off of the particle ID device.

momentum thresholds from 2.5, 9.0, 17.1 GeV/c to 0.6, 2.0, 3.8 GeV/c respectively as illustrated in Fig. 6.3. Shorter wavelength Cherenkov photons undergo Rayleigh scattering inside the aerogel itself. They are absorbed in the radiator or exit at random angles. A thin window between the aerogel and the gas radiator passes photons only in the visible range, diminishing the scattered component.

The same photo-detection system can detect Cherenkov rings produced in both the gaseous and the aerogel radiators if it operates in visible wavelengths.

6.3 RICH dimensions

The RICH detector can be located behind the tracking chambers just outside the central dipole magnet, about 4 meters away from the interaction point. The length of the RICH detector must be less than 3 meters to allow sufficient space for the EM calorimeter and the muon system. The aerogel radiator with thickness of about 4 cm will be mounted at the entrance to the RICH vessel. The gaseous C_4F_{10} radiator will fill the rest of the RICH volume. The transverse dimensions of the RICH entrance window are set by the aperture of the dipole magnet to be about 2.7 m $\times$ 2.7 m.

Spherical mirrors at the end of the gas volume reflect Cherenkov photons back and focus them into rings at the photo-detection surface. The photo-detectors should be located away from the flux of particles exiting the magnet, therefore the mirrors are tilted. In order to minimize geometric aberrations from an off-axis mirror configuration, the mirrors will be
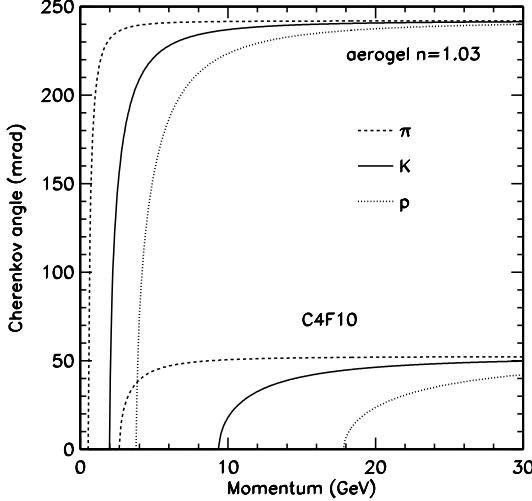
Figure 6.3: Cherenkov angles for various particle species as a function of particle momentum for C_4F_{10} and aerogel ($n = 1.03$) radiators.

split along the mid-line of the detector, reflecting photons to photo-detectors located on each side of the vessel in the non-bend view ($x - z$ plane). The longer the RICH detector the smaller the tilt angle. Since the geometric aberrations due to the mirror tilt are significant for the gas radiator, we plan on the longest RICH detector we can accommodate within the space limitations. This also maximizes the photon yield from the gaseous radiator, again improving the resolution of the device. Thus, the gas radiator will be approximately 3 meters long and the mirror radius will be 7 meters. The mirror tilt angle will be 240 mrad. Note that the mirror tilt angle cannot be further reduced by use of additional flat mirrors at the other end of the detector, like in the designs of HERA-B and downstream LHC-b RICH detectors, since the front of the gas volume must stay transparent to the aerogel photons. Such a scheme would be difficult even without aerogel radiator since there is not enough lateral space in the experimental hall. The transverse size of the mirror will be about 4 m × 4 m. A possible configuration for the mirrors is to make them from an array of individual hexagons. Each mirror half would consist of 18 full hexagons (76.2 cm tip-to-tip) and 6 half hexagons. A reflectance efficiency of about 90% has been achieved in HERA-B detector with 7 mm thick Pyrex coated with 200 nm Al and 30 nm MgF_2.

To find the size and optimal position and orientation of the photo-detection surface, we have used a ray tracing Monte Carlo. Even though the true focal plane of a spherical mirror is not planar, non-planar surfaces do not improve resolution significantly and are difficult to realize in practice. Thus, we have assumed that the Cherenkov rings are focused on a plane. Photo-detection systems considered (see the next section) work the best for normal light incidence. Thus the photo-detection plane must be tilted in the $x - z$ plane to follow the mirror tilt. The simulation indicates that tilt by 400 mrad produces a normal incidence on average.

Since the actual emission point along the track for Cherenkov photon is unknown, the Cherenkov angle reconstruction assumes emission at the track mid-point. The emission point error contributing to Cherenkov angle resolution is magnified by mirror tilt from 0.2 mrad to 0.51 mrad. This error imposed by geometrical considerations sets the scale for the other two major contributions to Cherenkov angle resolution: chromatic error and photo-detector segmentation error (called also photon position error) which can be controlled by parameters of the photo-detection devices.

The geometry of the proposed RICH detector is shown in Fig. 6.4.
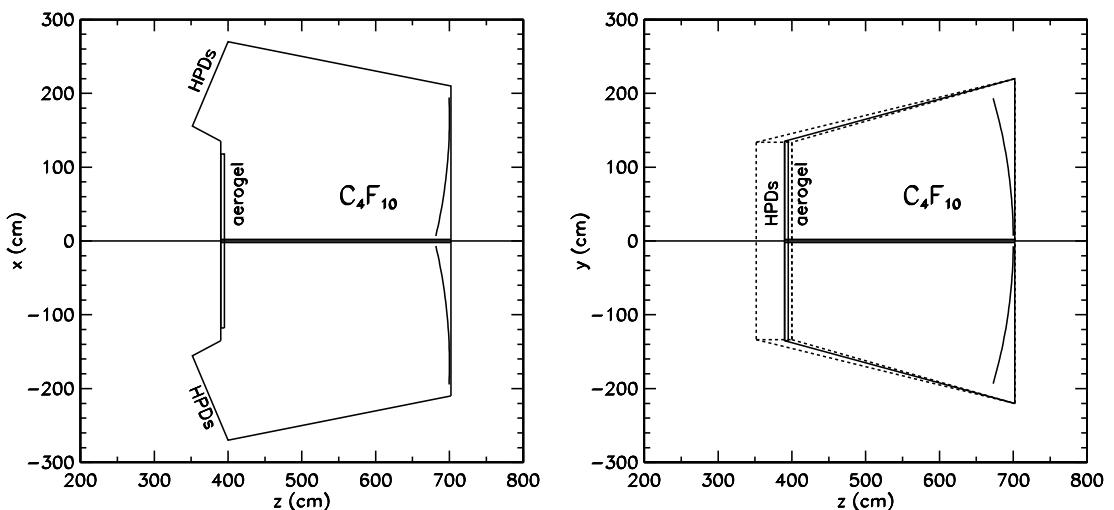
Figure 6.4: BTeV RICH geometry outline.

6.4 Photo-detectors

We choose to work in the "visible wavelength" regime, above 280 nm to minimize chromatic aberrations. Because of the open geometry of the forward spectrometer and the absence of a magnetic field around the RICH detector, arrays of photo-multipliers (PMT) or hybrid photo-diodes (HPD) can be used. We choose to use HPDs which offer a cheaper solution. Multi-anode PMTs are considered a back-up option. In this section we describe the HPD based system. The alternative PMT system is described and compared to our baseline choice in Appendix A.

HPDs are commercially available from DEP (Delft Electronic Products B.V.) in the Netherlands. DEP was founded in the late 1960s and is Europe's leading producer of a wide range of image intensifiers and photon counters (HPDs and MCPs) with markets in defense, medical and industrial sectors.

In an HPD, a photo-electron emitted by the photo-cathode is accelerated onto a seg-mented silicon diode by a very high voltage. Segmentation of the diode can be adjusted to

match specific needs for spatial resolution. Initial HPDs from DEP suffered from a large dead area around the diode. This has been greatly reduced by an electrostatic focusing system. DEP can now produce HPDs as large as 8.3 cm in diameter. These tubes were developed by DEP in collaboration with the LHC-b group. A drawing of a PP0380V tube from DEP is shown in Fig. 6.5. The tubes have circular cross-section.

A photon incident at a quartz window refracts and reaches a photo-cathode deposited on the inner window surface. The window has a spherical shape to provide some optical focusing of light onto the photo-cathode. About 80% of the tube area is active. Closely packed tubes will cover 91% of the area they occupy, thus overall geometrical light collection efficiency is 73%. Therefore, there is no need to use any additional demagnification systems to recover the geometric inefficiency.

In the HPD, a photo-electron emitted by a photo-cathode is directed and accelerated towards the diode. The photo-cathode is operated at a voltage of -20 kV with respect to the diode to provide enough energy to the electron for sizable charge generation in the silicon diode. The gain of the device is about 5000 and is proportional to the accelerating voltage. We expect to detect such signals with a signal-to-noise ratio of about 7:1. The charge collection from the silicon depends on the bias voltage and can be easily accomplished in a single 132 ns accelerator cycle. The present version PP0380V tubes contain a diode with 61 hexagonal pixels. The diode chip is bump-bonded to another chip with conductive traces arranging pixel outputs into a regular series appropriate for use with standard connectors. Pins attached to this chip take the signals out of the vacuum enclosure, where front-end electronics can be attached. Tests of the PP0380V tube by the LHC-b group showed that the electrostatic focusing works well and that good signal-to-noise can be obtained [6].

In the LHC-b RICH detectors, fine segmentation is required.[1] They plan to have 2048 pixels in each HPD tube, which makes it impossible to use the present DEP design in which individual pixel signals are taken to the outside of the tube. Therefore, the LHC-b group is developing their own silicon chip in which the diode is integrated with the front-end read-out and resides inside the tube.

The diode segmentation needed in BTeV is 163 pixels per tube. We can either follow the LHC-b path and develop our own integrated diode-readout chip to be put inside HPD by DEP, or we can rely on DEP and their subcontractors to modify their present scheme to accommodate more pixels per tube. Our current plan is the second solution that decouples development of read-out electronics from the manufacturing of the HPDs.

It appears that the IDE company from Norway already has a chip which can be easily modified to our read-out needs. Thus, no in-house chip development will be needed for the front-end processing. In the preliminary design we will have 163 hexagonal pixels per tube as illustrated in Fig. 6.5. The corresponding photon position error is 0.51 mrad. More

---

[1]The upstream LHC-b RICH detector is placed very close to the interaction point and therefore has small mirror radius, and consequently very compact Cherenkov rings. The downstream LHC-b RICH detector with large mirror radius is geared to identify high momentum tracks (BTeV does not need to cover these momenta because of the lower beam energy) and therefore a gas with low refractive index is used. This makes Cherenkov images compact again.
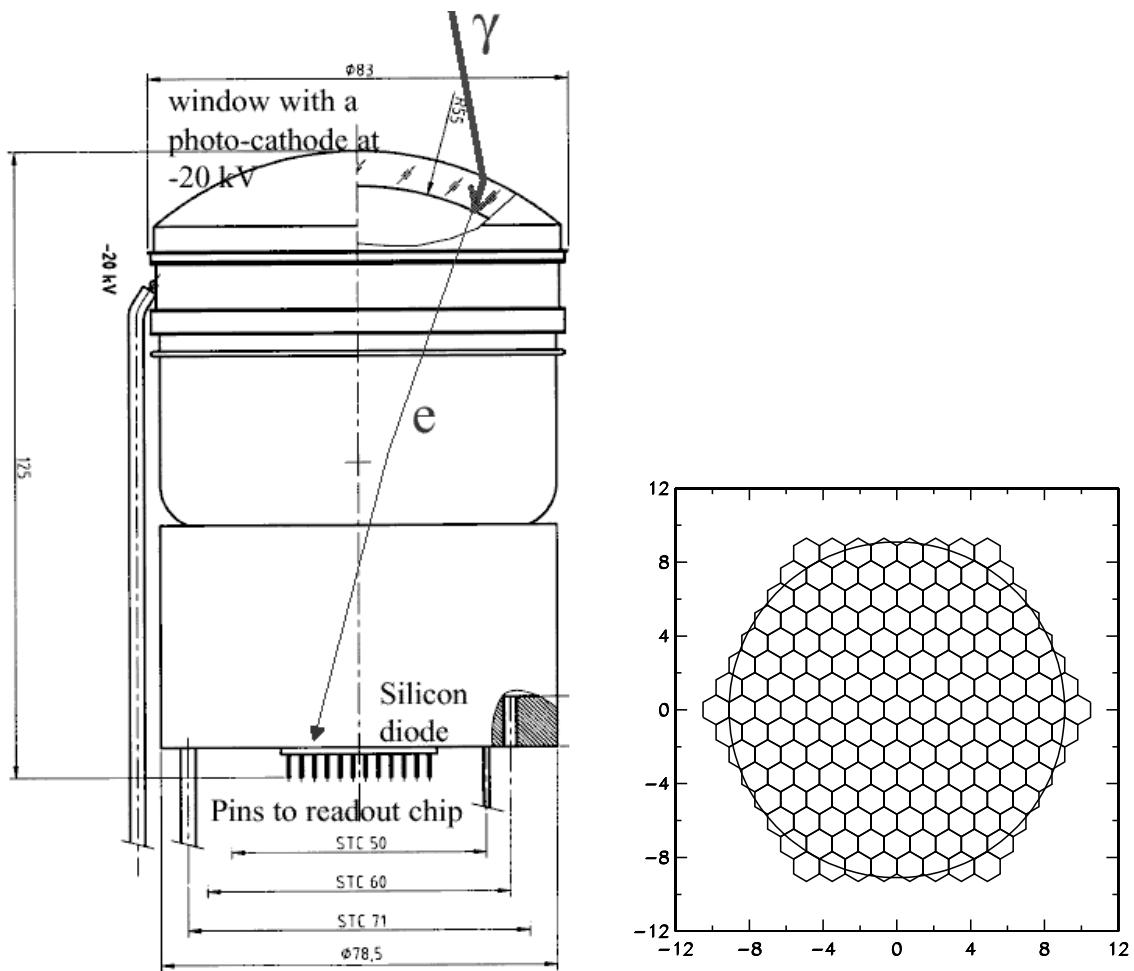
Figure 6.5: PP0380V HPD from DEP (on the left) and possible pixel structure on HPD silicon diode (on the right). Only parts inside the circular area would be actually implemented. Dimensions are expressed in mm.

information on the RICH read-out can be found in Appendix A.

The last major factor impacting the RICH performance is the wavelength coverage determined by the photo-cathode and window material. The wavelength sensitivity determines chromatic error (arising from wavelength dependence of the radiator refraction index) and is the major factor in the number of Cherenkov photons detected per track.

Quartz windows are a standard feature in the HPD tubes as they can easily sustain the large high voltage on the photo-cathode. High quality quartz extends the wavelength coverage from the visible range down to 160 nm. Such a large wavelength coverage results in a large chromatic error of 1.2 mrad per photon and in a large number of photons radiated per track (~ 235). When the wavelength coverage is limited, the photon yield drops but the chromatic error per photon improves. These two effects offset each other. The simulations show that a shallow optimum in Cherenkov resolution per track is reached when the wavelengths are limited to > 280 nm. This results in a chromatic error of 0.37 mrad per photon with a photon yield of 80 photons per track. Perhaps such wavelength coverage can be achieved with appropriate coating of the quartz window. Some R&D will be needed to resolve this issue. In the present simulations we assumed a bialkali photo-cathode with the 280 nm cut-off. The total Cherenkov angle resolution is 0.81 mrad per photon and 0.09 mrad per track.

To determine the number of HPDs needed, we plot in Fig 6.6 the $\pi^+\pi^-$ efficiency, for the requirement of 4σ K/π separation for both tracks, as a function of tube cost. For comparison, simulation of the PMT system is also shown. The system with 1888 HPDs (indicated by the vertical line in Fig 6.6) approaches the full geometrical coverage limit. We plan to order 192 spare HPD tubes which can be deployed and instrumented on the outside perimeter of the detector. With 163 pixels per tube, the entire detector (two identical RICH detectors) will have up to 339,040 electronic channels.

6.5 Expected resolution for the aerogel radiator

The chromatic error for aerogel photons is expected to be about 1.8 mrad. This is likely to be an overestimate since we did not simulate attenuation of shorter wavelengths in the radiator itself. The effect of the lucite filter between the aerogel and gas radiators was simulated as a sharp cut-off at 300 nm.

HERMES reports [3] an additional 0.7 mrad error, mostly due to aerogel density variation. The emission point error is expected to be small (0.2 mrad) because the radiator in only 4 cm thick. The detector granularity error is 0.6 mrad. Putting all these factors together we predict the Cherenkov resolution per photon to be around 2 mrad.

We assume 10 detected photo-electrons per track based on the recent reports from HERMES. This photon yield cannot be increased by using a thicker radiator because of light scattering and absorption in the aerogel itself. Further improvements in aerogel transparency would increase the number of photons per Cherenkov ring.

The expected Cherenkov resolution per track is about 0.6 mrad. Various contributions to the Cherenkov angle resolution for C_4F_{10} and aerogel radiators are summarized in Table 6.1.
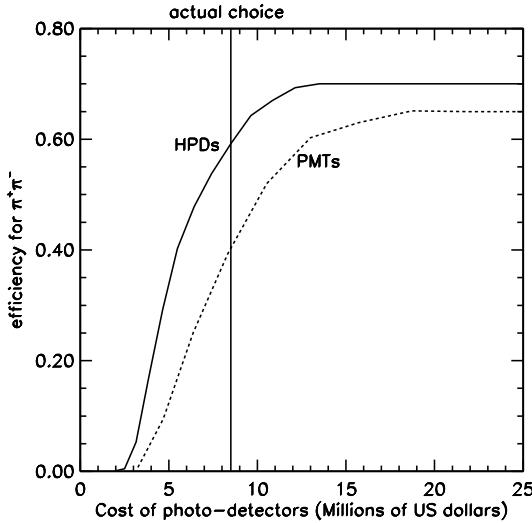
Figure 6.6: Efficiency for $B_d \to \pi^+\pi^-$ as a function of the cost of the photo-detectors. Efficiency is defined relative to the reconstructed events with both tracks crossing the RICH detector. Each track was required to have enough photons for 4σ K/π separation (from the gaseous radiator alone). Effects due to Cherenkov ring overlaps are not taken into account. The actual choice of a size of the HPD system is shown by vertical line.

More details related to the aerogel radiator can be found in Appendix A.

6.6 Expected physics performance

In a real environment, Cherenkov rings from different tracks overlap in the detector. This is illustrated in Fig. 6.7. A realistic simulation of efficiency and fake rates must take into account ambiguities in track-photon assignment.

We have analyzed simulated data with an algorithm which could be applied to real data. The reconstruction is performed in two steps. In the first pass, all hits within $\pm 3\sigma$ of a mass hypothesis are included in the per track average, excluding those hits which are within $\pm 3\sigma$ of the pion hypothesis for any other track. The second pass is essentially the same except that instead of assuming that all tracks are pions in the hit exclusion, the most likely mass hypothesis based on the first-pass results is used. To discriminate between two mass hypotheses for the same track (e.g. K or π) we cut on the modified χ^2 difference: $\Delta\chi^2_{K\pi} = \chi^2_\pi - \chi^2_K$, with $\chi^2_h = (\theta_{trk\,h} - \theta_h)^2/\sigma^2(\theta_{trk\,h}) - log(N_h)$. Here, $\theta_{trk\,h}$ (θ_h) is the measured (expected) Cherenkov angle per track for given mass hypothesis h, and N_h is the number of photons found within $\pm 3\sigma$ of this hypothesis. If no photons are found for a given hypothesis above the Cherenkov radiation threshold, a large χ^2 value is set. The best discrimination can be obtained by combining the χ^2 difference from the gaseous and aerogel radiators: $\Delta\chi^2_{K\pi} = \Delta\chi^2_{K\pi\,gas} + \Delta\chi^2_{K\pi\,aerogel}$. For a given cut value on the $\Delta\chi^2_{K\pi}$ we obtain a value for efficiency and fake rate.

89

Table 6.1: Expected performance of BTeV RICH system. The photon yield and the resolution per track given here do not take into account any reconstruction losses due to overlap of Cherenkov rings from different tracks in the same event.

	C_4F_{10}	Aerogel $n = 1.03$
emission point error	0.51 mrad	0.2 mrad
segmentation	0.51 mrad	0.6 mrad
chromatic error	0.37 mrad	1.9 mrad
total error per photon	0.81 mrad	2.0 mrad
number of photons	80	10
total error per track	0.09 mrad	0.6 mrad
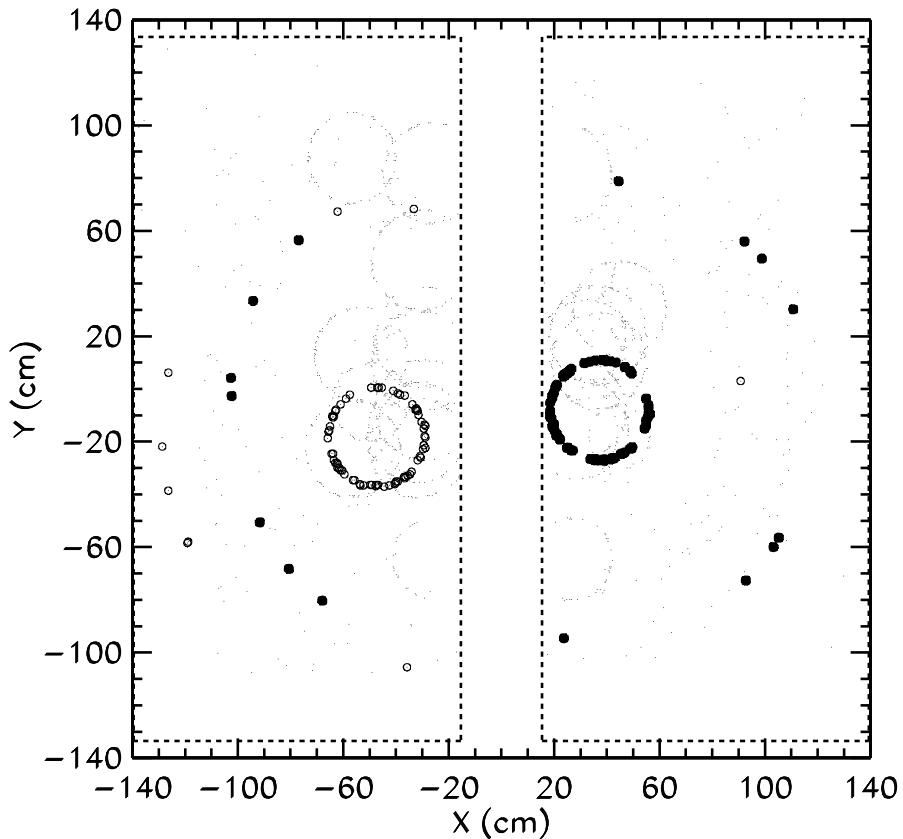
Figure 6.7: A typical reconstructed $B_d \rightarrow \pi^+\pi^-$ event. Gaseous and aerogel Cherenkov photons radiated by the signal pions are highlighted.
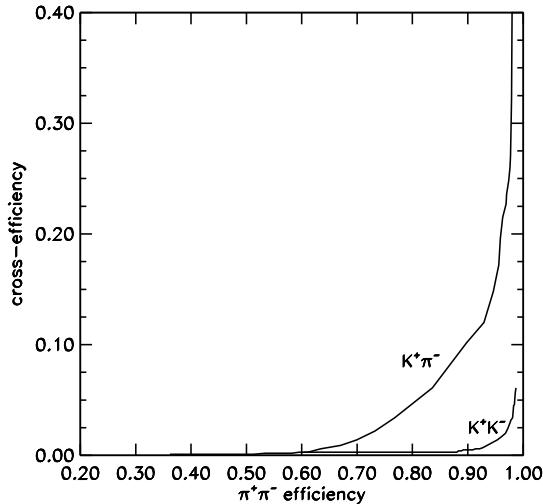
Figure 6.8: Cross-efficiency of particle identification system for $B_d \to K^{\pm}\pi^{\mp}$ and $B_s \to K^+K^-$ as a function $B_d \to \pi^+\pi^-$ particle identification efficiency. The efficiencies are defined relative to the number of events with both tracks entering the RICH detector. The Monte Carlo simulation included an average of two minimum bias interactions in addition to the $b\bar{b}$ production.

We have analyzed Monte Carlo samples of reconstructed $B_d \to \pi^+\pi^-$ (signal), $B_d \to K^{\pm}\pi^{\mp}$ (background) and $B_s \to K^+K^-$ (background) events. Background cross-efficiency is plotted as a function of $B_d \to \pi^+\pi^-$ efficiency in Fig. 6.8. Separation of $\pi^+\pi^-$ from K^+K^- is easier than from $K^{\pm}\pi^{\mp}$ as there are two tracks to discriminate on. The plot shows that the cross-efficiency can be made essentially zero for still sizable signal efficiency. A higher signal yield can be obtained if some background can be tolerated. The analysis method presented here can be improved by using the likelihood ratio for the two mass hypotheses, with the likelihood combining the χ^2 defined above with a probability for the observed number of photons found on the ring. Also, instead of removing photons from the ring overlaps, a global event fit of all mass assignments can be performed.

Bibliography

[1] W. Adam *et al.*, Nucl. Inst. & Meth. **A343**, 60 (1994).

[2] J.L. Rosen, Proceedings of The Fifth International Workshop on $B-$Physics at Hadron Machines, Los Angeles, USA, Oct. 13-17, 1997, published in Nucl. Inst. and Meth. A408 (1998) 191; S. Kopar, Proceedings of the The 3rd International Workshop on Ring Imaging Cherenkov Detectors Weizmann Institute of Science Ein-Gedi, Dead-Sea, Israel, Nov. 15-20, 1998.

[3] P. Carter, Proceedings of the The 3rd International Workshop on Ring Imaging Cherenkov Detectors Weizmann Institute of Science Ein-Gedi, Dead-Sea, Israel, Nov. 15-20, 1998. "Proposal for Dual Radiator RICH for HERMES", E. Cisbani *et al.*, Hermes Internal Note 97-005.

[4] LHC-B, Technical Proposal, CERN LHCC 98-4, LHCC/P4, 1998.

[5] R. Forty, CERN-PPE/96-176, Sept. 1996 published in Proc. of the 4th Int. Workshop on $B-$physics at Hadron Machines, Rome, Italy, June 1996, F. Ferroni, P. Schlein (Eds.), North-Holland, 1996.

[6] E. Albrecht *et al.*, "Performance of Hybrid Photon Detector Prototypes With 80% Active Area For The RICH Counters of LHCb", submitted to Nucl. Inst. & Meth. **A**, See also GLAS-PPE-1999-22 (hep-ex/0001053) and Nucl. Inst. & Meth. Phys. Res. **A433**, 159 (1999).

Chapter 7

Electromagnetic Calorimetry

7.1 Introduction

A thorough investigation of B decays requires the ability to study decay modes containing single photons, π^o's, and η's. The addition of a high quality crystal calorimeter to the CLEO detector was a first in coupling excellent photon detection to charged particle detection and demonstrated its importance in B decay studies [1]. Furthermore, the identification of electrons is most useful. Some of the important decay modes for BTeV include: $B^o \rightarrow (\rho\pi)^o \rightarrow \pi^+\pi^-\pi^o$, $B^o_s \rightarrow \psi\eta$, and $\psi\eta'$, and $B^o \rightarrow K^{*o}\gamma$ and $\rho^o\gamma$.

7.2 Choice of An Electromagnetic Calorimeter Based on PbWO$_4$ Crystals

Several technologies were considered that were radiation tolerant, had excellent energy and spatial resolution, were affordable and available. These included:

- A sampling calorimeter with lead and either scintillator or liquid argon

- A total ionizing calorimeter based on liquid krypton

- A calorimeter based on PbWO$_4$ crystals.

Lead and scintillator was rejected because of poor resolution. Fermilab is reluctant to deal with the numerous safety issues related to operating liquid argon or liquid krypton in the C0 enclosure. These devices also have long charge collection times which can, however, be overcome by careful pulse shaping. Liquid krypton, in particular, looked like an attractive possibility. Liquid detectors are radiation hard. However, the constant term in the energy resolution may be considerably larger due to the relatively large amount of material before the active sampling material [2].

We selected for our baseline a calorimeter made of PbWO$_4$ crystals. This technology has been developed by CMS. Our choice is based on several considerations:

- It satisfies our requirements on energy and spatial resolution. Blocks of size roughly 2.6 cm can be made. CMS expects to achieve a stochastic term of 2.7% and a constant term of 0.55% using Avalanche PhotoDiodes (APDs) for readout. We plan to use photomultiplier tubes (PMTs) and expect to have a smaller stochastic term, $\sim$1.6%.

- This material is very resistant to radiation damage, especially when doped with either Nb or La. CMS expects their crystals to survive a dose of $\sim$100 kGy. This should be sufficient for BTeV.

- They are fast – they deliver 99% of their light output within 100 ns, which is safely less than the bunch crossing time of 132 ns at the Tevatron.

The properties of $PbWO_4$ which are important for the calorimeter are given in Table 7.1.

Table 7.1: Properties of $PbWO_4$

Property	Value
Density (g/cm^3)	8.28
Radiation Length (cm)	0.89
Interaction Length (cm)	22.4
Light Decay Time (ns):	5(39%)
	15(60%)
	100(1%)
Refractive Index	2.30
Maximum of emission (nm)	440
Temperature Coefficient (%/^{o}C)	-2
Light output/NaI(Tl) (%)	1.3
Light output (pe/MeV into a 2" PMT)	10

7.3 Description of the BTeV Calorimeter

Our detector uses crystals very similar in size and shape to CMS. The CMS crystals are tapered to point at the interaction region. The crystals, which follow a 5 radiation length preradiator, are 220 mm long, and 26 mm square in the back [3]. Our crystals are the same length and virtually the same width. We also will use a projective geometry, but the taper is smaller.

The outer radius was chosen by weighting the efficiency of a larger calorimeter, up to the spectrometer acceptance of 300 mr, with the cost. In Fig. 7.1 we show the efficiency of detecting the reaction $\overline{B}^o \to D^{*+}\rho^-$, $D^{*+} \to \pi^+ D^o$, $D^o \to K^-\pi^+$ and $\rho^- \to \pi^-\pi^o$ as a function of calorimeter radius.

The calorimeter cost can be parameterized roughly as

$$T\ (\$) = 700(\$) \times N_c + 3,400,000(\$),\qquad(7.1)$$

Table 7.2: Properties of the BTeV electromagnetic Calorimeter

Property	Value
transverse block size, back tapered, smaller in front	26.0 mm × 26.0 mm 25.4 mm× 25.4 mm
Block length	22 cm
Radiation Lengths	25
Front end electronics	PMT
Digitization/readout	QIE (FNAL-KTEV)
Inner Dimension	± 9.88 cm × ± 9.88 cm
Outer Radius	160 cm
Total blocks per arm	11850

where N_c is the number of crystals and T the total cost. The fixed costs represent mainly startup of crystal production, the crystal container, the light calibration pulsar and electronics development. Weighing the efficiency versus total cost, we adopted an outer radius of 160 cm, corresponding to 23,700 crystals for both arms.

One end of the calorimeter is shown in Fig. 7.2 with the crystal hits displayed for an event generated with GEANT containing a $B^o \rightarrow \rho^o \pi^o$ decay. The two photons from the π^o decay are indicated by large circles. One photon has 19.3 GeV of energy, while the other has 2.4 GeV. The minimum energy displayed is 10 MeV. This corresponds to the minimum energy crystal we use in measuring the shower energy. It can be seen even from this one event that there is much more activity near the beam line than further out in radius. We will return to this point later.

CMS will use Avalanche Photodiodes (APD) to read out the barrel and Vacuum Phototriodes to read out the endcap. These devices are necessary because of the high magnetic field in which the crystals are positioned. In BTeV we are far from any magnetic fields, so we can use photomultiplier tubes. This provides less noise and better resolution as will be described below, in a well understood and tested technology. In fact, BTeV will use modified Hamamatsu R5380 tubes equipped with quartz windows to withstand radiation. These are similar to the ones used by KTeV.

As part of its acceptance criteria, CMS requires that all of their crystals emit more than 10 photoelectrons per MeV into a XP2262B photomultiplier tube covering the back face of the crystal. Using a smaller phototube, we expect to collect in excess of 7 photoelectrons per MeV. CMS actually collects ≈4 photoelectrons per MeV into two 25 mm^2 APD's on each crystal.

There are identical detectors in each arm of the spectrometer. Each detector is housed in a temperature and humidity controlled hut. There will be a dry nitrogen environment. Temperature stabilization is necessary because of the thermal coefficient of the $PbWO_4$ light output. In addition, the gains are monitored with a laser pulsing system.

Detectors are read out with PMTs. The light emitted by the crystal peaks at 440 nm. The output of the PMT will go to a modified QIE chip located near the photomultiplier
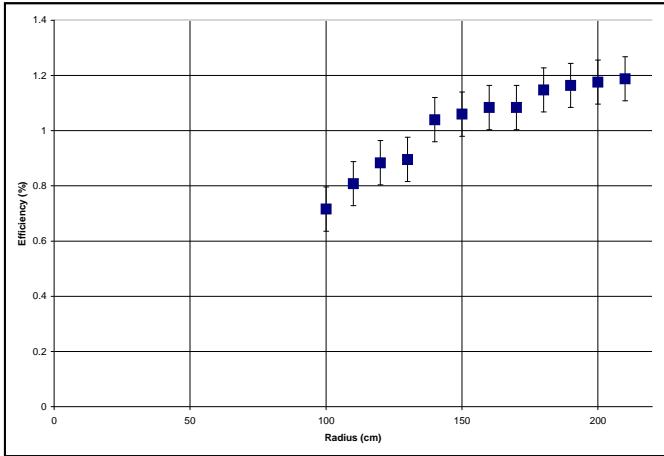
Figure 7.1: The efficiency as a function of calorimeter radius for the reaction $\overline{B}^o \rightarrow D^{*+}\rho^-$.

base [4], but outside the region of intense radiation. This chip provides a pulse height for each beam crossing. The expected light output is 7000 photoelectrons at 1 GeV. The detector is far enough away from the BTeV dipole so that there should be no need for special magnetic shielding from that source. Since we will magnetize the iron of the muon filter just downstream of the detector, we will put iron shield plates between the calorimeter's PMTs and the muon filter.

7.4 Effects of Radiation

The radiation levels at the crystals and phototubes are discussed in Appendix A. The maximum radiation levels occur close to the beam. There is also a relatively narrow vertical band of higher than average dose caused by the sweeping action of the BTeV dipole.

The simulations shown here were done for a smaller calorimeter of about 10,000 crystals. The integrated dose rates for most crystals are quite modest. In Fig. 7.3 (left) we show the dose distribution in the crystals (here dose means the maximum dose in any part of the crystals), for 1 year of running at a luminosity of $2 \times 10^{32} \text{cm}^{-2}\text{s}^{-1}$. We also show (right) the cumulative fraction of crystals with doses less than that shown on the horizontal axis. We see that ~90% of the crystals have a yearly accumulated dose of less than 1000 Gy (0.1 Mrad).
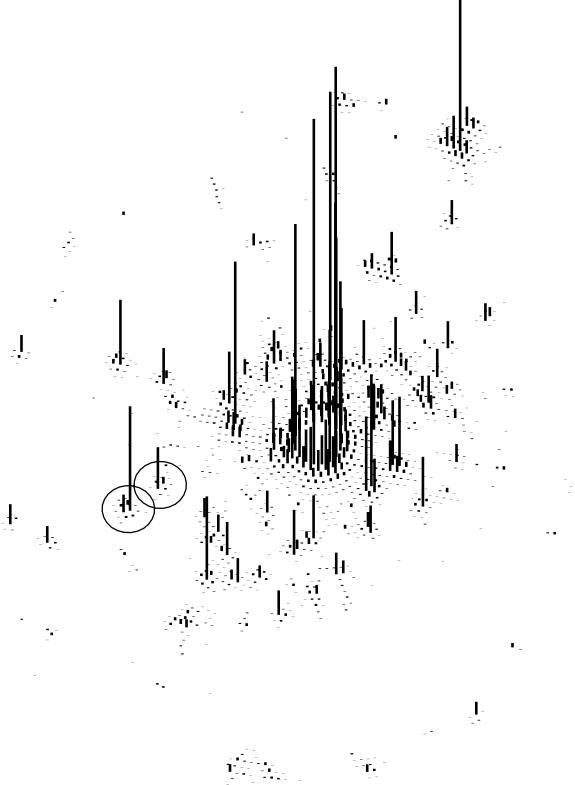
Figure 7.2: The energies in the PbWO$_4$ calorimeter (one end) for an event containing two photons from the decay sequence $B^o \rightarrow \rho^o\pi^o$, $\pi^o \rightarrow \gamma\gamma$. The photons of energies 19.3 and 2.4 GeV are surrounded by large circles. All energies above 10 MeV are shown, with the height of line proportional to the energy.

7.5   Expected Resolution

The detailed estimates of the expected resolution are given in Appendix A. Briefly, 22 cm long crystals were selected after a GEANT study determined that length to be optimal. The transverse size was chosen to be essentially the same as CMS is already using, though making the crystals one to two mm larger, if possible, would sacrifice little performance and cut the costs by 4-8%, due to a reduced number of photomultiplier tubes and electronics channels.

Using 7 photoelectrons/MeV we find contributions to the stochastic term in the energy resolution to be $1.2\%/\sqrt{E}$ from photon statistics and $1.08\%/\sqrt{E}$ due to crystal size and clustering (transverse and longitudinal leakage) where E is in units of GeV. For the constant term we use the CMS estimate of 0.55%. We note that KTeV has achieved a constant term
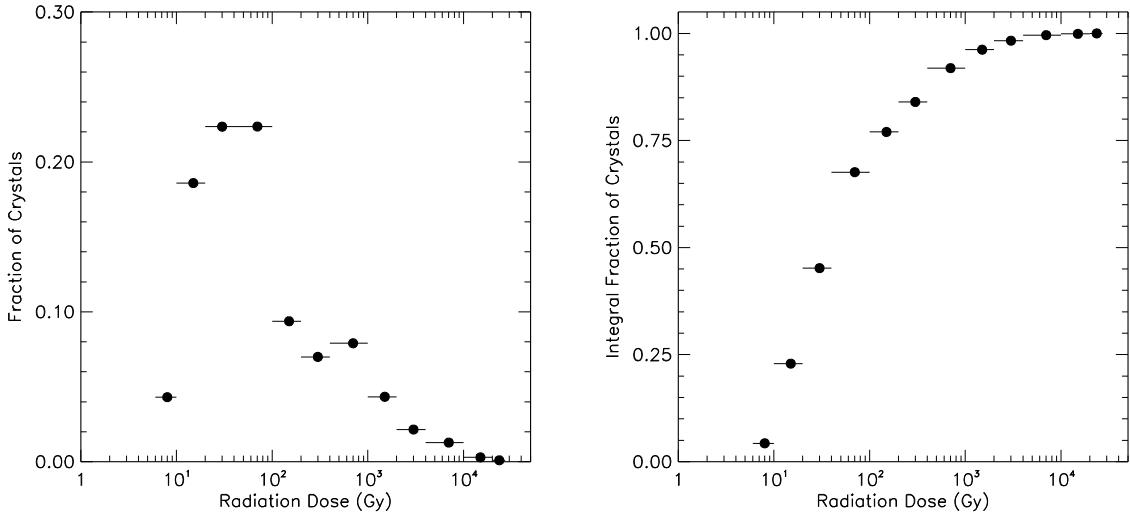
Figure 7.3: (left) The maximum dose distribution in the crystals for for 1 year of running at a luminosity of $2 \times 10^{32}\text{cm}^{-2}\text{s}^{-1}$. (right) The integral of the maximum dose distribution.

of 0.45% [5]. Overall we expect the energy resolution to be

$$\frac{\sigma_E}{E} = \sqrt{\frac{a^2}{E} + b^2} = \frac{a}{\sqrt{E}} \oplus b, = \frac{1.6\%}{\sqrt{E}} \oplus 0.55\% \qquad (7.2)$$

The spatial resolution in both directions transverse to the crystal axis is expected to be

$$\sigma_x = \frac{3500 \ \mu m}{\sqrt{E}} \oplus 200 \ \mu m \ . \qquad (7.3)$$

With these resolutions, the mass resolution is excellent. Fig. 7.4 shows the invariant $\gamma\gamma$ mass for π^o's of 10 GeV energy incident on the calorimeter where the energies and positions are simulated by GEANT and our cluster finder is used to reconstruct the photons. The mass resolution is 2.6 MeV, which compares favorably to that found in CLEO (5 MeV) [6].

7.6 Examples: Photon Efficiencies in $B^o \rightarrow K^*\gamma$ and $B^{\rightarrow}\rho\pi$

In the hadron environment, there is great concern that there is so much activity from charged track interactions in the calorimeter and additional photons that the signal photons will be totally obfuscated. As a test case we consider the decay $B^o \rightarrow K^*\gamma$. Although we are only concerned with the gamma reconstruction here, we require that the two charged tracks from the K^* decay reach the RICH detector, in order to insure that the B^o's were in the
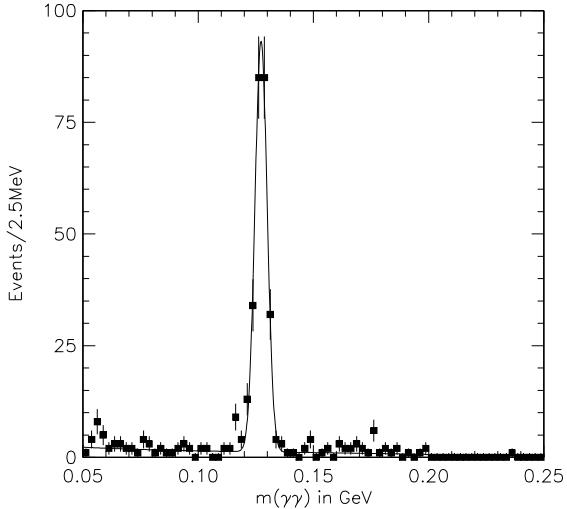
Figure 7.4: The $\gamma\gamma$ invariant mass for 10 GeV π^o's incident on the calorimeter. The fit is to a Gaussian signal function plus a polynomial background.

acceptance of the spectrometer. The decay was simulated by GEANT at a mean interaction rate of two per crossing.

To identify photons we find clusters of energy whose shape is consistent with an electromagnetic shower and we require that no charge track extrapolate close to the cluster. In Fig. 7.5 we show the difference between measured and generated photon energy divided by the generated energy.

The BTeV resolution is about a factor of two better than CLEO. The CLEO spectrum is absolutely clean; there is a small background in the BTeV distribution. Our concern is a large overlap with fragments from other particles that would cause the photon energies to be shifted high and out of the peak, but this apparently is not the case. In Fig. 7.6 we show radial distribution of generated photons from $K^*\gamma$, and accepted ones and the efficiency, where we define accepted photons as ones with energies within 3σ of the peak that pass the shower shape and isolation cuts. We note that the same number for the CLEO barrel calorimeter calculated in the same manner using GEANT is 89%. While our efficiences start out considerably lower than CLEO, they increase rapidly and demonstrate the usefulness of the calorimeter.

We also looked at the π^o efficiency as a function of the radius of the π^o at the z position of the calorimeter for $B \to \rho^o\pi^o$ and $B \to \rho^+\pi^-$ decays. The efficiency, shown in Fig. 7.7 plateaus at a radial distance of about 75 cm. This simulation was run in a larger than proposed calorimeter so we could view the dependence on radius more easily.
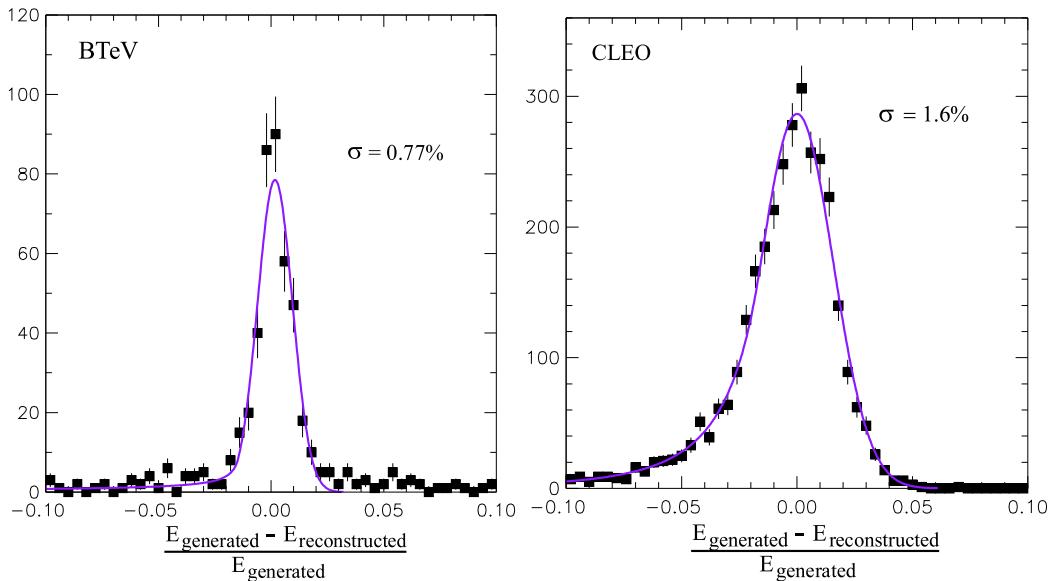
Figure 7.5: The difference between the measured and generated energies, divided by the generated energy, for reconstructed photons as simulated by GEANT for BTeV (left) and CLEO (right). The photons candidate clusters were required to have shower energy shapes consistent with that expected for photons and to be isolated from charged tracks. The BTeV simulation was run at 2 interactions/crossing.

7.7 Crystal Production

There are two producers of PbWO$_4$ crystals working for CMS. One of these companies is in Russia, the Bogoroditsk Techno Chemical Plant, and the other is in China at two sites, Beijing and Shanghai. CMS crystal production is scheduled to finish in 2005. We have visited the crystal production facilities in Russia and in China. Our Russian and Chinese collaborators have been most helpful in setting up these visits. The Russians already have been producing production crystals (>6000) and are eager to have our business. They have supplied prices and possible schedules. The Chinese have not started production as of this writing, but they are very close to doing so. They are also very interested in BTeV production. Our Chinese collaborators at Shandong University are also interested and capable of producing PbWO$_4$ crystals. We would like to initiate a startup program with them as soon as possible. Generally we think it important to have more than one supplier of crystals. Because of the open nature of the BTeV detector, crystal installation can proceed in place over a long period of time.
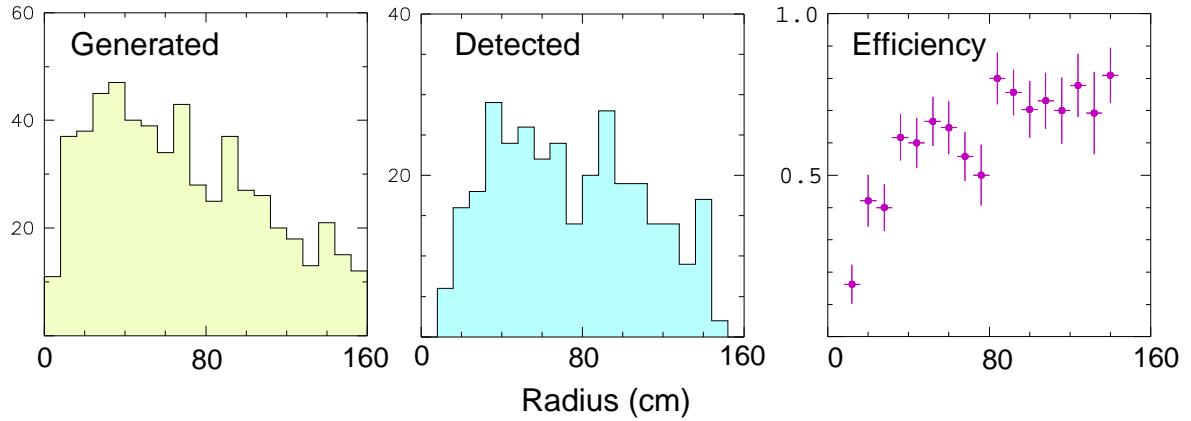
Figure 7.6: The radial distribution of generated and detected photons from $B^o \to K^*\gamma$ and the resulting efficiency. The detector was simulated by GEANT and the resulting crystal energies were clustered by our software. The charged tracks from the K^* were required to hit the RICH. The simulation was run at 2 interactions/crossing.
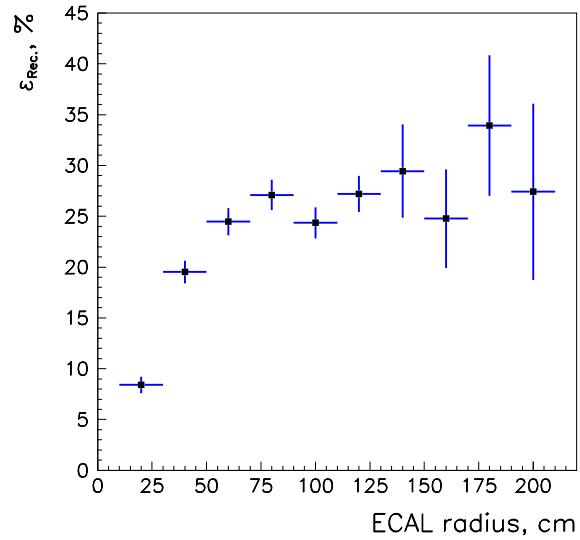
Figure 7.7: The efficiency as a function of the radial distance from the beam line of π^o's from $B \to \rho\pi$ decays at the position of the calorimeter. The simulation was run at 2 interactions/crossing in the "large" calorimeter.

Bibliography

[1] E. Blucher, B. Gittelman, B. K. Heltsley, J. Kandaswamy, R. Kowalewski, Y. Kubota, N. Mistry, A. Bean, and S. Stone, "Tests of Cesium Iodide Crystals for an Electromagnetic Calorimeter," Nuclear Instruments & Methods **A249**, 201 (1986).

[2] We note that a past favorite, CsI (undoped), was attractive from many points of view. However, it is not radiation hard enough. Furthermore, it is not quite fast enough to completely produce all its charge in under one crossing. Lead glass has been rejected because it is not likely to survive the yearly radiation dose.

[3] CMS, "The Electromagnetic Calorimeter Project Technical Design Report," CERN/LHCC 97-33, CMS TDR 4 (1997).

[4] R. J. Yarema *et al.*, A Fast Wide Range Charge Integrator and Encoder ASIC For Photomultiplier Tubes," FEMILAB-PUB-92-311 (1992). This chip was developed for KTEV.

[5] Private communication from Ed Blucher. The dominant cause of the constant term seems to be a differing crystal response depending on where on the face of the crystal the incident photon strikes. This effect was not observed in the CLEO CsI calorimeter (see ref. 1).

[6] Y. Kubota *et al.*, "The CLEO II Detector," Nuclear Instruments & Methods **A320**, 66 (1991).

Chapter 8

Muon Detector

The BTeV muon system has two primary functions:

- J/ψ and Prompt Muon Trigger: Besides providing interesting physics (including J/ψ final states of B decays, direct J/ψ production, and semileptonic decays), this trigger performs an important service role by selecting (bottom) events on which the more aggressive and technically challenging vertex trigger can be debugged and evaluated.

- Muon Identification: Many of the experiment's physics goals (rare decay searches, CP violation studies which require tagging, studies of beauty mixing, searches for charm mixing, etc.) rely on efficient muon identification with excellent background rejection.

We have selected a toroidal magnet design combined with fine-grained tracking elements. This design permits a "stand-alone" trigger: *i.e.* a muon trigger based solely on information from the muon detector. In addition, improved background rejection is possible by comparing this measurement with momentum and tracking information from the rest of the spectrometer. The system design has been chosen to reduce and uniformly distribute occupancies and to minimize confusion in pattern recognition. To provide a viable trigger, the system must obtain a rejection rate at Level 1 of a few hundred. The goal for muon misidentification is 10^{-3}.

Given the objective of a stand-alone trigger and the size limitations set by the experimental hall, one can make fairly general calculations that place specific (and restrictive) constraints on the design of the system. We first describe these calculations and use them to motivate the overall design of the muon system. We then describe the trigger efficiency and rejection studies we have performed.

8.1   General Design Considerations

The fractional momentum resolution in a magnetic spectrometer can be parameterized as $\sigma_p/p = \sqrt{a^2 + (b\,p)^2}$ where the a term depends on the bending power and multiple scattering environment of the detectors and the b term depends on the bending power and the detector
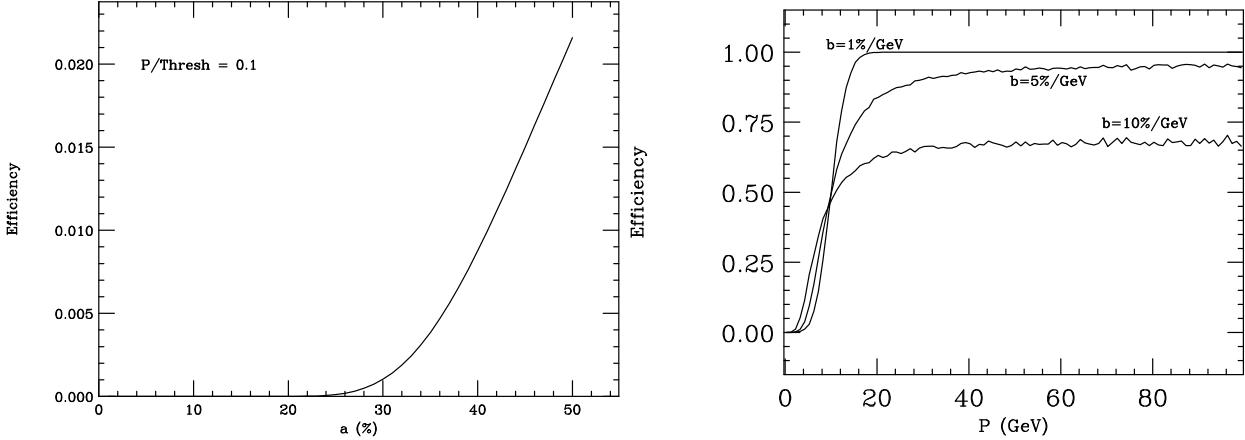
Figure 8.1: (left) Efficiency, calculated in a simple Gaussian model, for a muon with 1/10 the threshold momentum to pass a momentum threshold as a function of a, the MCS term in the resolution formula. (right) Efficiency as a function of muon momentum for a trigger designed to fire with a 50% efficiency at 10 GeV. The multiple scattering dominated term is fixed ($a = 25\%$) while the measurement dominated term varies from $1\%/\text{GeV}<b<10\%/\text{GeV}$.

layout and spatial resolution. Figure 8.1a shows that the potential rejection provided by a trigger for low momentum muons significantly degrades once the low momentum fractional resolution exceeds 25%. At $a = 25\%$, the trigger rejects very soft muons at roughly the 4σ level. At higher momentum, where multiple scattering is less important, one becomes sensitive to the b term. The high momentum resolution influences how sharp a momentum threshold one can make in a stand-alone muon trigger. Figure 8.1b illustrates this point by showing the trigger efficiency as a function of momentum for several b values. These considerations suggest minimum performance criteria of $a < 25\%$ and $b < 1\%/GeV$.

8.2 Baseline Muon System

Several measurement and shielding scenarios were studied before reaching the baseline design. In assessing possible layouts, we compute the momentum resolution using an error matrix which incorporates Gaussian models for the detector resolution and multiple Coulomb scattering. The interaction region is modeled as a Gaussian beam spot with $\sigma_x = \sigma_y = 1$ mm and $\sigma_z = 30$ cm. The result of these studies gives us the baseline geometry shown in Fig. 8.2.

8.2.1 Baseline Geometry

Two toroids, 1 m long with 1.5 T fields, provide the bending power. The muon detectors will be set up in three stations, one between the toroids and two behind the toroids. The momentum can be measured using the two, well shielded, downstream stations and the nominal beam constraint. The station between the two toroids provides a powerful confirming hit to eliminate fake tracks.
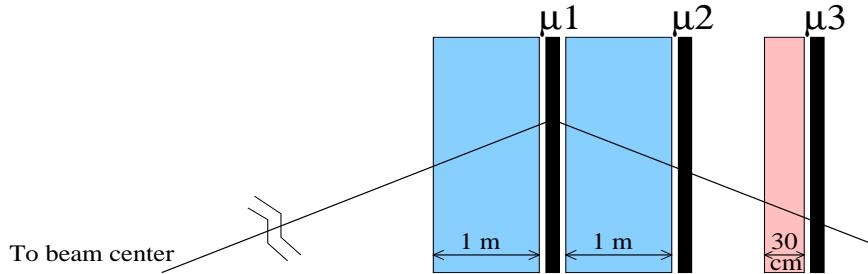
Figure 8.2: Layout of the baseline geometry, shown in elevation view. The three dark boxes, labelled μ_1, μ_2, and μ_3, represent detector stations with 4 measurement views per station. The two lighter boxes with lengths of 1 m represent magnetized steel toroids, which provide bending power for the muon momentum measurement and which also serve as hadron absorbers. The 30 cm long lighter box is an unmagnetized iron shield. The downstream trajectory is measured by μ_2 and μ_3. The upstream trajectory is measured using the nominal beam center with possible help from μ_1. To obtain sufficient bending power, both 1 m steel sections must be magnetized.

8.2.2 Baseline Detector

The basic building block in the construction of a detector station is a "plank" of 3/8" diameter stainless steel proportional tubes as shown in Fig. 8.3. Thirty-two tubes are arranged in a double layer with an offset of half a tube ("picket fence" geometry) and will be soldered at each end to a brass gas manifold and supported in the middle by soldering to brass support rib piece(s). This provides a sturdy, self-supporting building block which acts as an excellent Faraday cage. Proportional tubes have been selected as the detector technology because they are robust and have the necessary rate capability. We intend to use a fast gas (*e.g.* 88% Ar, 10% CF_4, 2% CO_2 ($v_d \approx 9$ cm/μs) [1]) so the maximum collection time (drift plus charge integration) for a signal should be less than 60 ns, allowing us to gate off hits due to incoming beams which arrive 70 ns before particles from the interaction region. We will use thin walled (0.01") stainless steel tubes. This is a proven technology — it has been used successfully by the CDF collaboration where similar tubes [2] had a long lifetime with low failure rate. The tubes will be strung with 30 μm gold-plated tungsten wire.

The 0.5 cm wire spacing of this design has no dead regions and has an effective spatial resolution of 5 mm/$\sqrt{12}$ = 1.4 mm which meets our requirements outlined in Section 8.1. Figure 8.4 shows the momentum resolution for various muon system configurations assuming a 2.5 mm resolution and incorporating the magnetic fields and multiple scattering. The top curves show the result for only one magnetized toroid which is clearly ineffective. The bottom curves show results when both steel filters are magnetized. The case where only a beam constraint and μ_2–μ_3 are used is illustrated by the dashed red curves. Including information from μ_1 produces the solid magenta curves. Curves are shown for three different azimuths. The chosen geometry exhibits good performance relative to our criteria on a and b. The importance of μ_1 lies in providing redundant information to eliminate fake tracks and for matching tracks with the inner tracker at higher trigger levels and offline. Its effect on
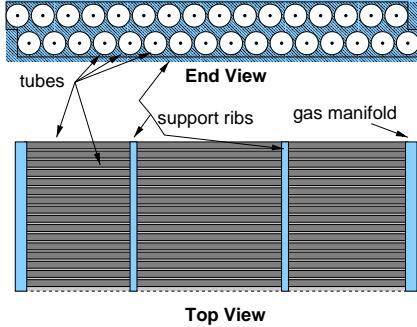
Figure 8.3: End and top views of one "plank" of proportional tubes.

Table 8.1: Parameters of the baseline BTeV Muon System.

Radial coverage	16–240 cm
Toroid Z-location (center)	± 830, 960 cm
Average Station Z-location	±890, 1020, 1150 cm
Total Length	3.3 m
Toroid Length (each)	1 m
Toroidal Field	1.5 T
Tube cell size	1 cm (diameter)
Wire spacing:	0.5 cm (staggered)
Spatial resolution	1.5 mm
Total channels	39,936 per arm
Momentum resolution	$\sigma_p/p = 19\% \oplus 0.6\% \times p$

the momentum resolution is less important.

To minimize occupancy at small radii and to minimize pattern recognition confusion, each detector station will consist of eight overlapping pie shaped "octants," as shown in Fig. 8.5a. The four views (r, u, v, and r) in each octant are shown in Fig. 8.5b. The r (radial) view is repeated to provide redundancy for the most important view and to help reject fake tracks in the trigger. The u and v views are rotated $\pm 22.5^{\mathrm{o}}$ from the r view and are used to measure ϕ to resolve hit ambiguities, thereby reducing the misidentification rate. The views stack on top of each other and are built from the planks described above. There will be 13 planks in each view of an octant. Pairs of octants will be combined into quads which will be the structure moved in and out of the BTeV detector.

A summary of the baseline BTeV muon system is given in Table 8.1. Further details may be found in Appendix A.

8.2.3 Front-end Electronics

Although we will only be reading out a latch bit and not using TDC's, we are considering gating the system to exclude hits coming from the incoming beam. Therefore the front-end
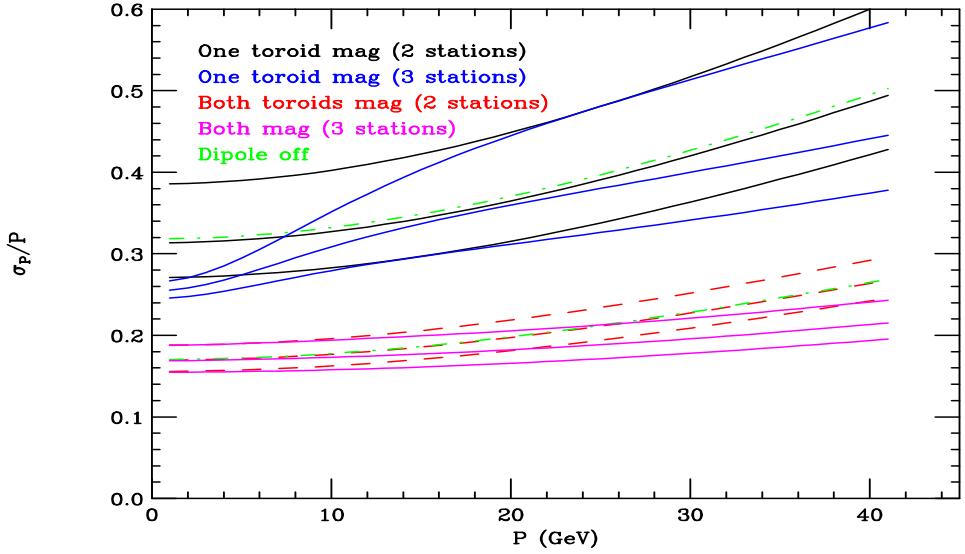
Figure 8.4: Plots of the fractional momentum resolution versus momentum for various muon system configurations. All plots assume 2.5 mm RMS spatial resolution. The plots appear in groups of 3 corresponding to $\phi = 270^o$ (best) 0° and 90° (worst). The ϕ dependence illustrates interference between the central dipole and muon toroid(s). The top (bottom) curves show the resolution when one (two) toroids are magnetized. The black and red curves illustrate the case where only the two stations after the second toroid are used. In this case the trajectory upstream of the toroid comes entirely from the beam constraint. The blue and magenta curves illustrate the case where the station between the two toroids is used to help determine the trajectory prior to the toroids. The green curves assume only the final two stations are used, and the dipole is turned off.
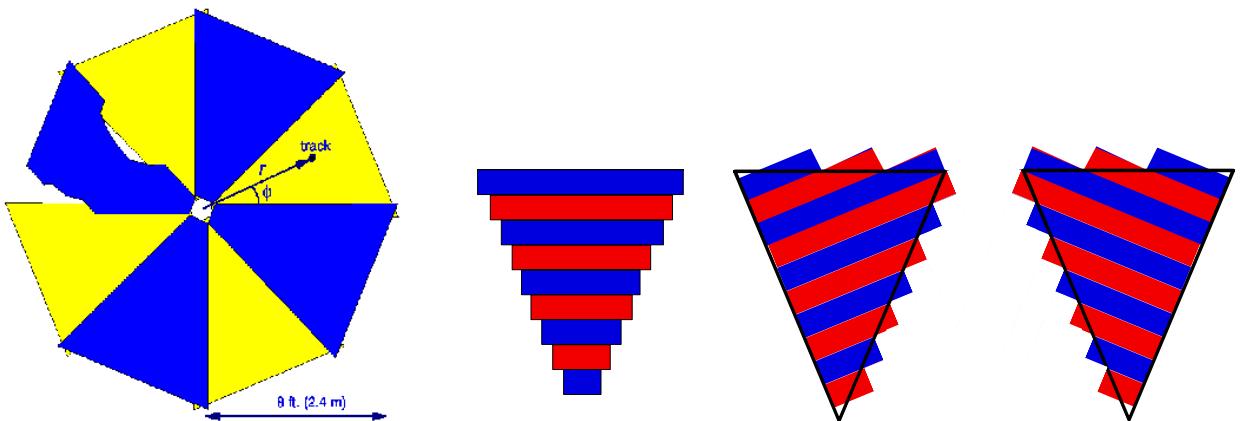
Figure 8.5: (left) Beams-eye view of one muon detector station, which consists of eight overlapping octants arranged in two layers. One octant is cut away in places to show the overlap between adjacent octants. (right) Arrangement of planks to form the four views in an octant (r view is repeated). There will be 13 planks per view (more than shown).

107

electronics need a double pulse resolution < 30 ns. The electronics should also be low cost and have sufficient gain. We intend to mimic the setup used for the CDF COT. There will be a PC board to deliver high voltage to each proportional tube, and a PC board with electronics to amplify and digitize the tube signal. Both boards will be located directly at the end of a plank.

We plan to utilize the ASDQ integrated circuit developed at the University of Pennsylvania to amplify and digitize the signals coming from the proportional tubes. This chip will be used in the Run-II CDF central outer tracker for a similar purpose. The ASDQ amplifies the first ~8–10 ns of the signal and outputs an LVDS (equivalent) differential logic signal. This chip, when mounted on a PC board, has a low effective threshold of about 2 fC and features a double pulse resolution of ~20 ns. The chip seems to be an ideal choice for the BTeV muon system. Beam tests of the predecessor to the ASDQ, the ASD8B, indicate that it is suitable for our use. Noise problems encountered during the beam test with the ASD8B required creating a Faraday cage out of Aluminum, copper plated G10 and copper tape. We will address this problem by completely enclosing the electronics and proportional tube connections in a Faraday enclosure. This is the reason that the gas manifolds (Fig. 8.3) will be made of brass and soldered to the proportional tubes.

The ASDQ digital signals will be sparsified, serialized, and read out using a standard Fermilab readout protocol. Fiber optic cables will transfer the data from the plank to a buffer memory. Slow control and monitoring functions will be performed via fiber optic link as well.

Additional details regarding the front-end electronics and readout can be found in Appendix A.

8.3   Trigger Studies

We have studied the triggering performance of our baseline muon system using both MCFast and full GEANT simulations. We used minimum bias events to study rejection rates and $B^0 \to J/\psi K_s^0$ and $B_s^0 \to J/\psi K^{*0}$ where $J/\psi \to \mu\mu$, to investigate trigger efficiency. The minimum bias events were generated with PYTHIA and include elastic scattering, single and double diffractive, low p_T scattering and semi-hard QCD $2 \to 2$ processes. The number of events per crossing is generated from a Poisson distribution with average two. In generating the signal events, a Poisson distributed number of minimum bias events (average of 2) were added to the signal event.

We studied the performance of the baseline muon system using several levels of simulation. The first of these simulation studies was designed to study the acceptance for $B^0 \to J/\psi K_s^0$ events and rejection of minimum bias events using the MCFast simulation package. This study allowed us to make detailed studies of the kinematic cuts required to create an effective trigger. The MCFast simulations were based on the parameterized Gaussian momenta resolutions which were obtained using the semi-analytic calculations shown in Fig. 8.4.

While the MCFast simulations were quite useful for initial testing of muon triggers, to have confidence in our ability to reject minimum bias background we require more realistic

simulations. This was achieved using a full GEANT simulation which simulates additional hits in the muon system due to non-prompt sources, δ-rays, electromagnetic shower debris and hadronic shower leakage. After describing this simulation we discuss properties of the "noise hits" and conclude by describing the performance of several (stand-alone) muon triggers within this realistic environment.

8.3.1 GEANT simulation description

The geometry used in the GEANT simulation is similar to our baseline geometry. There are two 1 m long cylindrical toroids with inner radius of 14 cm and outer radius 240 cm. One muon station is placed between the two toroids and two behind the toroids. As described in Section 8.3.2 it was found that additional shielding in front of the third station and around the beam pipe dramatically improved the noise levels so this was added. The aluminum support structure was also simulated. The tube walls were simulated in a simplified form using a planar sheet and in an exact form using cylindrical tubes. In this simulation, each octant had one r-view, one u-view, and one v-view. We expect that the extra downstream r-view can only improve the results. Therefore, this simulation is a pessimistic prediction.

All default processes plus δ-rays were simulated. Photons and electrons were traced down to an energy of 0.1 MeV while hadrons and muons were traced down to 10 MeV.

The digitization was performed in two, alternate, ways. In the approximate tube wall simulation (with planar sheets) a sensor volume composed of a large sheet was used. When this sensor volume was reached, the coordinates were used to determine the tube which should be fired. The efficiency for this method was 100% for tracks within 95% of the tube radius. In the simulation which used cylindrical tube walls, an individual sensor volume equal to 85% of the radius of the tube was used inside each tube. Therefore, knowing which sensor volume a particle was in gave the tube number directly. The information which was output for the trigger included the type, momentum, and parent of each particle which left a hit in the muon system as well as a list of all hits associated with the particle.

Using computer farms at the University of Puerto Rico–Mayaguez and Vanderbilt University we generated over 100,000 minimum bias events and more than 10,000 B to J/ψ final state signal events. Events were simulated assuming two minimum bias interactions/crossing.

8.3.2 Properties of additional "noise" hits

Figure 8.6 shows the characteristics of the hits in the muon detector for $B_s^0 \rightarrow J/\psi \, K_s^0$ events with the original geometry. The noise hits are dominated by low momentum secondaries coming from interactions with the walls of the beam hole in the muon filters as is evident from the radial position distributions at the downstream faces of the filters. Secondaries spray out of the downstream ends of the holes in both filters. Figure 8.7a documents this problem. The hit distribution in the muon planes can be understood by considering that the percentage of such tracks striking a given plane will increase with the lever arm between the nearest filter and the plane. (All planes go down to the same radial position.) The

effect is noticeable in stations 1 and 2 which are just downstream of filters but is largest for station 3. All planes in station 3 receive a large flux due to the long lever arm between this station and the second filter. To reduce this effect we added a 30 cm thick filter with the same radial coverage as the toroids just in front of the third station. The improvement is presented in Fig. 8.7b. Interactions with the beam pipe as well as δ-ray production also contribute significantly to the noise. Additional shielding (8 cm) around the beam pipe is also proposed and the resulting improvement is shown in Fig. 8.7c and 8.7d.
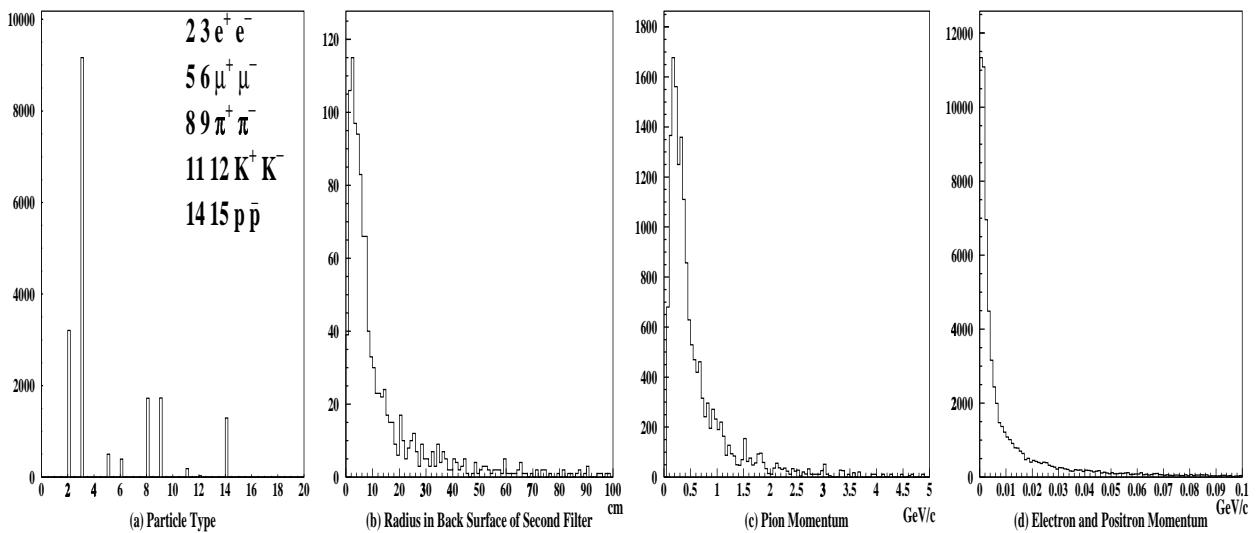
Figure 8.6: Characteristics of muon detector hits for $J/\psi\,K_s^0$ events. (a) Particle type. Each particle type is assigned an integer code, shown in the legend. For example, π^+ is given the code 8. The electromagnetic and hadronic noise is larger than the muon signal. The excess of electrons over positrons is due to δ-ray production. (b) Projected radial position at the downstream face of the second filter for tracks that hit station 3. The large percentage of tracks emanating from the vicinity of the hole in the filter is quite evident. (c) and (d) Momentum of pion and e^+/e^- noise. Noise secondaries have much lower momentum than J/ψ muons.

8.3.3 A muon "tracking" trigger

To establish an "upper" limit on muon trigger performance in the presence of the GEANT generated noise we studied the performance of a muon "tracking" trigger. This trigger loops over all hits within a given octant sector to choose the set of hits which forms the best muon track with the nominal beam center. The best set of hits is defined using a χ^2 test to the hypothesis that the muon system hits form a track which emanates from the nominal beam origin with production angles x'_o, y'_o, and momentum p. It is important to emphasize that this is a simulation of a "stand-alone" trigger. There is no use of any tracking information from the inner tracker; all information comes from the muon system.
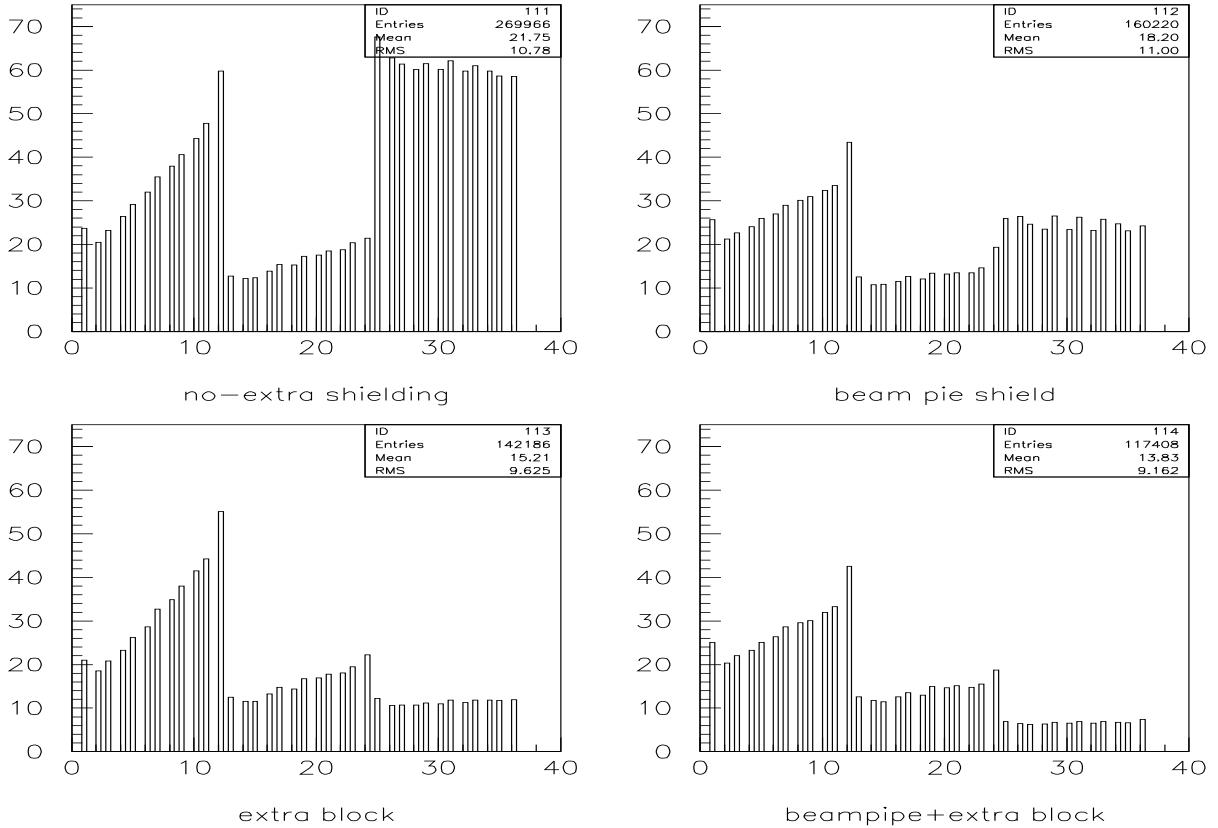
Figure 8.7: Distribution of hits among muon detector planes for different shielding configurations. There are twelve planes per station. (a: upper left) Original design; no extra shielding. (b: lower left) Extra shielding block in front of the third station. (c: Upper right) Extra shielding around the beam pipe. (d: Lower right) Both beam pipe and third station shielded.

For expediency we make several simplifications. We use a least-squares χ^2 — $e.\,g.$ there is no attempt to build the χ^2 including the correlated covariance matrices due to multiple Coulomb scattering. We also use a significantly simplified magnetic description. A sudden bend approximation is used for the central dipole field. The toroid field in each octant is modeled by a constant field pointing perpendicular to the octant sector bisector and is thus an approximation to a circular field. Particles are traced through the toroids according to the sudden bend approximation. The tracks are assumed to pass exactly through the nominal beam center and no attempt is made to incorporate the smearing of the nominal beam center in constructing the χ^2. There is no attempt to take into account dE/dx loss. All equations are linearized in the 3 fit parameters x'_o, y'_o, and q/P which means that the fit is a classic, non-iterative linear fit.

It is important to note that although the fitting process has been simplified, the GEANT modeling of the muon system has not. A complete magnetic trace is used throughout. Appropriate multiple scattering and dE/dx losses are incorporated. A realistic luminous

111

region is used and non-prompt muons are generated from detached vertices.

Figure 8.8 quantifies the rejection power and relative efficiency of this tracking dimuon trigger. The J/ψ efficiency is plotted versus the minimum bias rejection ratio. The J/ψ efficiency is normalized to events where both muons from the J/ψ have momentum greater than 5 GeV/c and both muons also leave hits in all three stations of the muon system. All reconstructed muons are required to have $\chi^2 < 25.2$ and the two muons must have opposite reconstructed charge and appear in two different octants. The main branch of the "cut tree" is a requirement on the minimum radius of all the muon hits. Branching off is a set of increasingly tighter cuts on the maximum χ^2. The third branch is a set of cuts on the minimum reconstructed P_T.

We are able to achieve a rejection in excess of 500 to 1 with an efficiency of nearly 50% with the principal cuts being a radius greater than 32 cm, a $\chi^2 < 14.4$, and a minimum $P_T > 0.4$ GeV/c, although several different cut selections give essentially identical results.
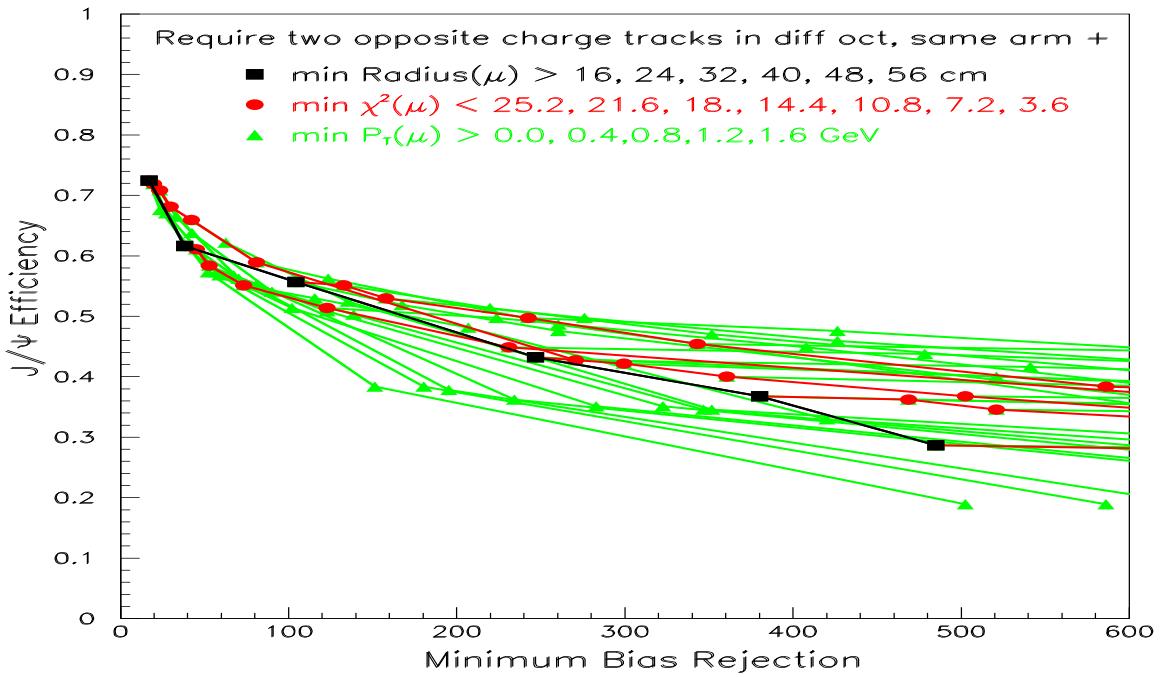
Figure 8.8: J/ψ efficiency versus minimum bias rejection rate for cuts on the minimum radius, maximum χ^2, and the minimum muon P_T. Two oppositely charged tracks in different octants in the same arm must pass the relevant cut. Black lines connect the squares which show the variation as the minimum radius is increased. From each black square a red line connecting the circles shows the change as the maximum χ^2 cut is decreased. From each red circle a green line connecting the triangles shows the variation as the minimum reconstructed muon P_T is increased. All information in this stand-alone trigger comes from hits in the muon system alone.

8.3.4 Simplified Tracking Trigger

We have also studied a simpler trigger which only uses the r-view hits and ignores the dipole contribution. It uses a more efficient algorithm that only fits likely candidate hits. Results for this trigger are shown in Fig. 8.9 where the J/ψ efficiency is plotted versus the minimum bias rejection ratio. The cut variables include the reconstructed momentum and p_T. In addition, the minimum radius cut requires that all hits associated with the track are at a radius greater than the radius cut. This shows that a very simple stand-alone dimuon trigger can provide rejection rates up to 600:1 while retaining an efficiency of 25%.
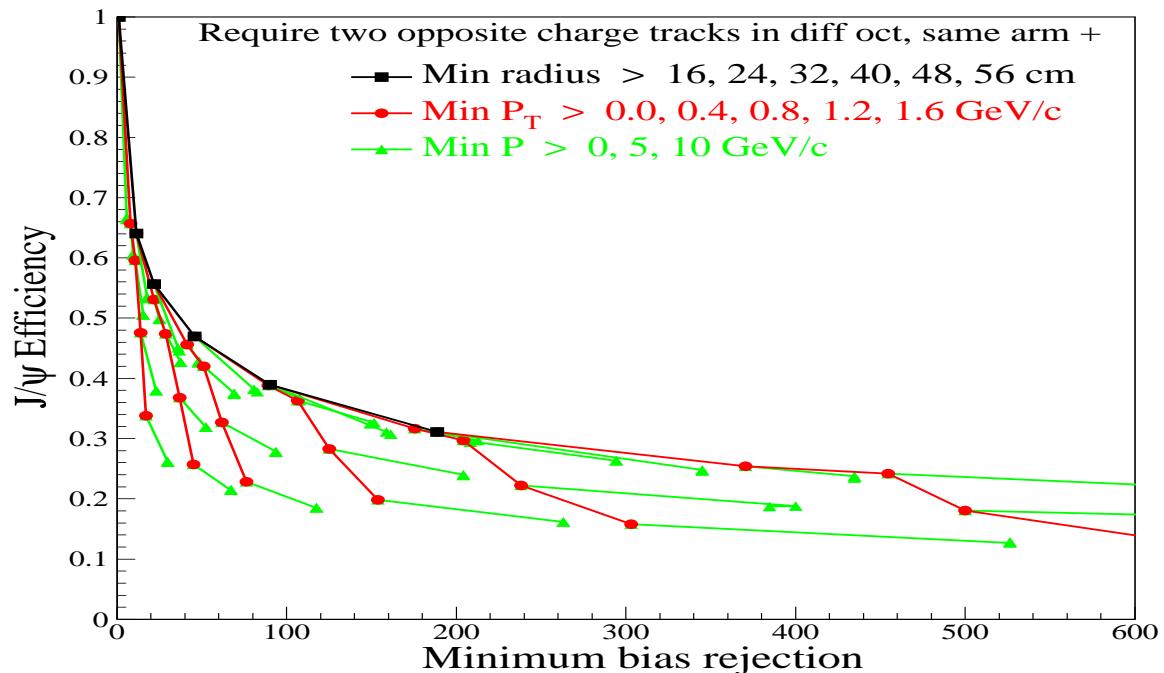
Figure 8.9: J/ψ efficiency versus minimum bias rejection rate, using a simplified tracking trigger, for cuts on the minimum radius, minimum muon p_T, and the minimum muon momentum. Two oppositely charged tracks in different octants in the same arm must pass the relevant cut. Black lines connect the squares which show the variation as the minimum radius is increased.. From each black square a red line connecting the circles shows the change as the minimum muon p_T cut is increased. From each red circle a green line connecting the triangles shows the variation as the minimum muon momentum is increased.

8.3.5 Trigger Summary

Our original MCFast simulation only included muon hits due to real muons. Therefore, the entire minimum bias background came from pion and kaon decays-in-flight. This produced an average of 1.2 muon tracks in a minimum bias event with a resulting average occupancy of 0.015%. From this we were able to obtain rejection ratios of 600 to 1 with efficiencies of 85%. Using a much more realistic GEANT simulation (described in Section 8.3.1) brings

the average occupancy up to 1.2% for minimum bias events (2 per crossing) and 2.1% for B to J/ψ events (plus 2 minimum bias events). Using a straightforward tracking algorithm described in Section 8.3.3 we are still able to retain a 600:1 rejection ratio with efficiency of nearly 50% even though our occupancy has increased by a factor of 100. This trigger will be implemented using a modification of the silicon pixel vertex trigger. If this proves too difficult, we have the option to use the simpler trigger described in Section 8.3.4 which also achieves a 600:1 rejection ratio at a still acceptable efficiency of 25%. These efficiencies and rejection rates were calculated using a simulation which did not include the second r-view. Therefore, we expect that these results will only get better. The dimuon trigger, which rejects at 600:1, uses a small fraction of the Level 1 bandwidth which is dominated by the vertex trigger which rejects at 100:1. Therefore, this trigger is suitable for both calibrating the vertex trigger and taking physics data. Following the Level 1 trigger (at either Level 2 or Level 3), we can gain further rejection by requiring the tracks which are found by the muon system to correspond to tracks found by the silicon pixels and forward trackers. After requiring a link we can also cut on the more accurate forward-tracker momentum, require the track be detached from the primary vertex, and/or make an invariant mass cut.

Bibliography

[1] J.M. Butler *et al.*, "Study Of Fast Gases, Resolutions And Contaminants In The D0 Muon System," Nucl. Instrum. Meth. **A290**, 122 (1990).

[2] S. Bhadra, S. Errede, L. Fishback, H. Keutelian and P. Schlabach, "The Design And Construction of the CDF Central Drift Tube Array," Nucl. Instr. Meth. **A268**, 92 (1988).

Chapter 9

The BTeV Trigger

In this chapter, we begin with a description of the trigger requirements for BTeV. We then discuss the strategy used to achieve these goals and provide some details of the hardware implementation and associated trigger algorithms. We conclude with a summary of the trigger performance. Additional details of the trigger algorithm and performance are presented in the simulation section, and a more detailed discussion of the trigger hardware is provided in Appendix A.

9.1 Goals and Requirements for the BTeV Trigger System

BTeV plans to study a broad range of B decays and to employ many different B tagging techniques. Other experiments [1] include a fairly simple "first level" of triggering which, while significantly reducing the number of events so that the higher trigger levels have more time to deal with the survivors, also restricts the types of final states that are accepted. Trigger strategies that require the presence of specific final-state particles, such as muons, or demand the presence of a few high-p_t hadrons, are examples of this. BTeV, instead, will trigger on the key property that differentiates B (and charm) particles from other particles, namely their characteristic lifetimes. This allows our trigger to reject light-quark background events at the necessary $\sim 10^{-3}$ level and still be efficient for beauty and charm decays containing only hadrons in the final state (e.g. $B_s \to D_s K$, $B^o \to \rho\pi$ and charm states such as $D^o \to K^+K^-$), as well as those containing leptons (such as $B^o \to \psi K_s$ or $B_s \to \psi\eta$).

This strategy requires the BTeV trigger to do track and vertex reconstruction on every beam crossing and to search for evidence of an interaction followed by the decay of a particle within a few hundred microns to a few millimeters away from the primary vertex. In practice, this is done by reconstructing all primary vertices and selecting events that have additional tracks with large impact parameters with respect to the nearest primary vertex.

116

9.1.1 Rejection Requirement

The goal is to reduce the full interaction rate of 15.2 MHz at a luminosity of 2×10^{32} cm^{-2}s^{-1} (7.6 MHz beam-crossing rate with an average of 2 interactions per crossing) to about 4 kHz of crossings written to permanent storage. The rate of events containing B decays with daughter particles that all fall inside the BTeV acceptance is about 1 kHz. We expect that the trigger will not be perfect, since some B events will have only some of their tracks within the acceptance but will still satisfy the trigger, and since some charm events will also satisfy the trigger. Therefore, we allow a safety factor of 4. It follows that the trigger must accept no more than approximately one out of every 2000 beam crossings.

9.1.2 Efficiency Requirement

To obtain clean samples of heavy-quark events, decay-vertex cuts must be made in the offline analysis. A vertex-based trigger maximizes our trigger efficiency, because the events we reject with online vertex cuts have a strong correlation to those that would eventually be rejected during analysis. The efficiency requirement for the trigger is that it must accept $\approx 50\%$ of the events that would have survived in an ideal world in which the full offline analysis could have been run on every beam crossing.

9.1.3 Rate and Deadtime Requirement

Since the beam crossing rate is 7.6 MHz, a trigger decision must be made on average every 132 ns. We want the deadtime from all sources to be practically zero: deadtime $\ll 10\%$.

9.2 BTeV Trigger Strategy

In order to make a trigger decision on average in 132 ns, we plan to employ a three-level trigger scheme (Fig. 9.1). Each level has the same goal: to select events with detached vertices typical of heavy-quark decays. The levels differ in the parts of the BTeV detector whose data they have access to, the amount of time on average they have to make their decision, and the type of hardware used to implement them.

Level 1 uses information from the pixel detector only. It finds and fits tracks, constructs primary vertices, and looks for tracks with large impact parameters signifying a B decay. To speed up the calculation, Level 1 finds tracks only as they enter the pixel detector and as they exit it. Momentum matching is used in linking the entering and exiting track segments together into tracks. Since the pixel detector is embedded in a 1.6 T magnetic field and has spatial resolution better than $10 \, \mu$m, the track momentum can be crudely measured even within a track segment that spans only three pixel stations. After segment linking, an improved momentum estimate is used to weight tracks in vertex fits, and in the calculation of normalized miss distances.

Detector

| 7.6 MHz crossing rate | **Buffer** | 200 KB/ Event | **Level 1** (secondary vertex, muon)
Interaction rate: 15.2 million/sec
Latency: variable (average time 0.15 ms)
Reduction: ~ 100 |

Accept / Reject

| 70 − 100 kHz | **Buffer** | 200 KB/ Event | **Level 2** (refined tracking and vertex cut)
Partial reconstruction
Latency: variable (average time 20 ms)
Reduction: ~ 5 − 10 |

Accept / Reject

| 15 − 20 kHz | **Buffer** | 200 KB/ Event | **Level 3** (complete events)
Partial and full reconstruction
Latency: variable (average time 200 ms)
Reduction: ~ 5 − 10 |

2 − 4 kHz

Accept / Reject

Data Quality Monitor

Data Storage 40 KB/ Event

Figure 9.1: Overview of the three-level trigger system.

The Level 1 trigger operates at the full beam-crossing rate and is the most demanding part of the trigger. It is able to carry out track reconstruction at this high rate because of the very high-quality, low-noise, three-dimensional tracking information provided by the pixel detector. Our specifications for the maximum allowable data-transfer bandwidth out of Level 1 and into Level 2, and for the maximum computing power available at Level 2, require that Level 1 reject $\sim 99\%$ of all beam crossings that do not contain B events. The trigger efficiency, as defined above, should be at least 60%.

Level 2 redoes the pixel-detector reconstruction, but this time finds all the hits associated with each track found at Level 1. It then combines information from the pixel detector with information from the first three stations of the forward tracker. The extra information allows Level 2 to reject fake or poorly-measured tracks, and to calculate the momentum of each track with much better precision. This enables it to do fits with a good estimate of the

covariance matrix for the tracks. It then searches for tracks with large impact parameters with respect to the primary vertex. Since Level 2 operates on the 1% of all beam crossings that pass Level 1, on average it must make a trigger decision every $13.2\,\mu$s. We expect that Level 2 will reject 90% of all crossings that pass Level 1, while retaining more than 90% of crossings containing a B event.

Level 3 provides the last factor of 2 or 3 of rejection. It does full pattern recognition using the pixel detector and all stations of the forward tracker. It then performs a full three-dimensional vertex reconstruction to find primary, secondary, and tertiary vertices. Based on this vertex information (plus some specialized cuts to rescue a few difficult-to-trigger topologies such as single-prong secondaries), events are selected for permanent storage. Level 3 is expected to be nearly 100% efficient on crossings containing B events. Since it only needs to inspect one crossing out of 1000, it has on average 132 μs to make a trigger decision.

To reduce the amount of data recorded, Level 3 can also summarize the raw event to some extent. Especially at the highest luminosities, only part of the information of the event will be written out. For example, the calorimeter, RICH, and muon information on the side away from the B candidate will be discarded.

The computations required by the Level 1 algorithm take much longer than $132\,$ns even with the very fast processors that will be available when the experiment runs. We deal with this by implementing parallel pipelines, so that the computations are done in parallel by many processors. There is thus a long latency between the time the event occurs and the times of the various trigger decisions. The parallelism entails handling many crossings at once, as well as analyzing the data from each crossing with some degree of sub-event parallelism. The system uses a very large number of advanced commercial components — mainly field-programmable gate arrays (FPGAs) and digital signal processors (DSPs) — to achieve the required throughput.

Data must be retained for the time it takes to make the trigger decision. This is done in a very large buffer memory. Since this memory does not need to be in the radiation field near the detector, inexpensive mass-produced computer memory can be used. To reduce the amount of data that needs to be stored while the trigger decision is being made, the data for each crossing are "sparsified" or "zero-suppressed" as they come off the detector (via the front-end electronics) into the buffer memory. We estimate that an average crossing will have $200-300\,$kbytes of sparsified data. We plan to have total buffer memory of ≈ 0.5-1.0 terabyte, which is significantly more than required to handle this and all other levels of buffering.

Since computations will be occurring for many (actually a few thousand) crossings simultaneously, and the computations for a single crossing will be carried out on more than one processor, it is a challenge to keep all processors fully occupied and not waiting for information from other processors. To achieve the best pipeline operation, we do not require trigger decisions to be time ordered. It is "normal" for trigger decisions to be made for some crossings while processors are still working on much earlier crossings. We have tried to reduce the number of synchronization points, where idle time could occur, to a minimum. We have provided buffering wherever such a problem could appear. As soon as a Level 1

decision is made, a Level 1 accept or reject is sent back to the buffer-memory managers so that the buffers can be freed for new events (in the case of a reject) or retained for the Level 2 trigger (in the case of an accept). The Level 2 trigger can reject an event at any of several points in its algorithm, as can Level 3. As soon as the trigger decision is known, the answer, either a reject or an accept, is sent back to the buffer-memory managers. Time-outs will be employed in various places throughout the system to prevent "thrashing" on pathological events and to handle certain failures, for example the "death" of a processing node.

9.3   Trigger System Implementation

9.3.1   Trigger System Overview

Fig. 9.1 shows the conceptual scheme we envision, with Level 1 running at the full crossing rate. The Level 1 vertex trigger is the main physics trigger, but the ancillary dimuon triggers are simple to implement and (by allowing a sample of heavy-quark events to be recorded with no vertex-reconstruction bias) will provide a cross-check of the efficiency of the vertex trigger. Events satisfying Level 1 will be sent to a farm of Level 2 processors, which will be general-purpose CPUs. The Level 2 processors will request event data from additional detectors, mainly from the forward tracker and the muon detector, in order to refine the trigger decision, with still more event data requested as successive cuts are passed. Once a complete event has been built, the event is by definition at Level 3. At Level 3, a stripped-down version of the offline analysis can be performed, and events can be compressed into a summary format to reduce the needed bandwidth to record events for permanent storage.

Although separate buffers are shown in Fig. 9.1, the Level 1, 2, and 3 buffers will most likely be implemented as a single physical buffer, *i.e.* , the pieces of the event will reside in the Level 1 buffers for the individual detector subsystems until called forth by each level of the trigger algorithm. This arrangement (also used by CMS [2]) serves to reduce the event-building bandwidth, since the event is built up gradually as it progresses to higher and higher trigger levels, with complete events assembled only at the Level 2 output rate. Fig. 9.2 is a more detailed block diagram showing these features.

For definiteness we have based our design studies on hardware that is now (or soon to be) available. We recognize that rapid progress in the electronics and computing industries means that when the actual design choices are made (≈ 2004), many of these parts — microprocessors, digital signal processors (DSPs), field-programmable gate arrays (FPGAs), random-access memories (RAMs), and so forth — will have been superseded by devices having greater performance per dollar.

It is important to recognize that although the trigger is a technical challenge, it is based on chips, processors, memories, and network components that are commercially available, with speed and capabilities that are steadily growing, and with prices steadily dropping.
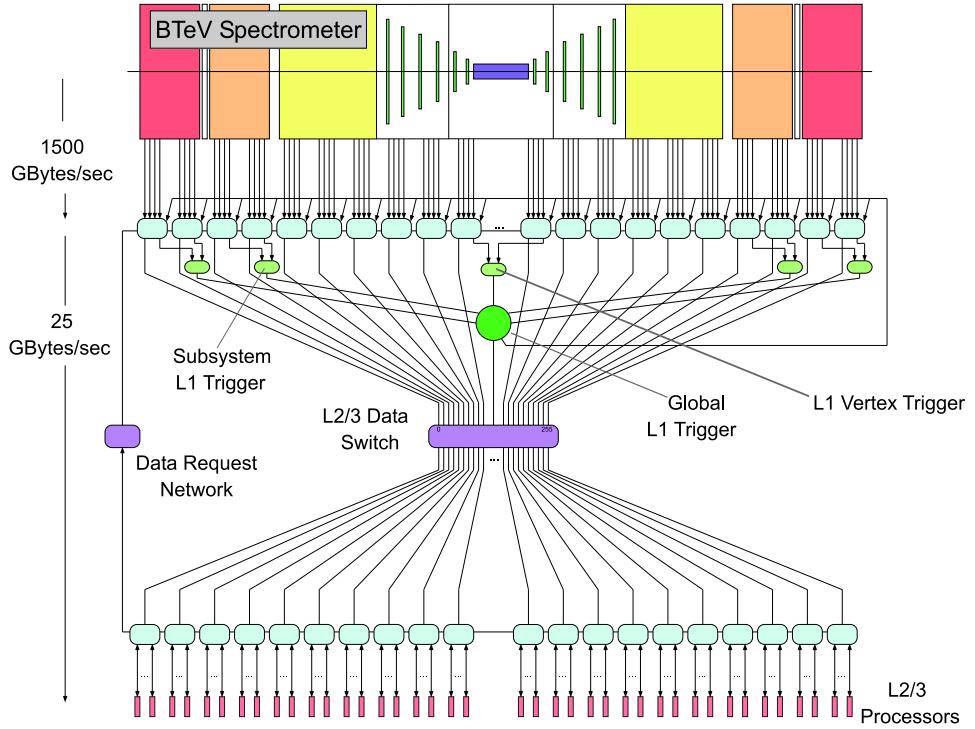
Figure 9.2: Block diagram of proposed DAQ system indicating design bandwidths before and after the Level 1 trigger: up to 1500 GBytes/s of data are digitized from detector subsystems and stored in Level 1 buffers, and up to 25 GBytes/s are passed to Levels 2 and 3.

9.3.2 Level 1 Vertex Trigger

The "centerpiece" of the BTeV trigger system is the Level 1 detached-vertex trigger (Fig. 9.3). We have considered various track-finding algorithms, and while the one we discuss here (called "BB33") already meets our requirements, we may yet revise it to lower costs or improve performance. The BB33 algorithm seeks to reduce combinatorial processing time by 1) use of track-finding hardware featuring massive parallelism and 2) careful choice of the region of interest for track "seeds."

The Level 1 algorithm has three phases:

- Track "segment finding";

- Segment matching and track fitting; and

- Vertex finding.

The most demanding part of the baseline algorithm (in terms of both computing power and data-transfer bandwidth) is identifying track segments in each set of three adjacent stations ("segment finding"). This is done in a custom hardware processor using FPGAs. The track segments are then linked into tracks using farms of DSPs. Because of the criticality of the segment finder, we provide a rather detailed discussion below.

121
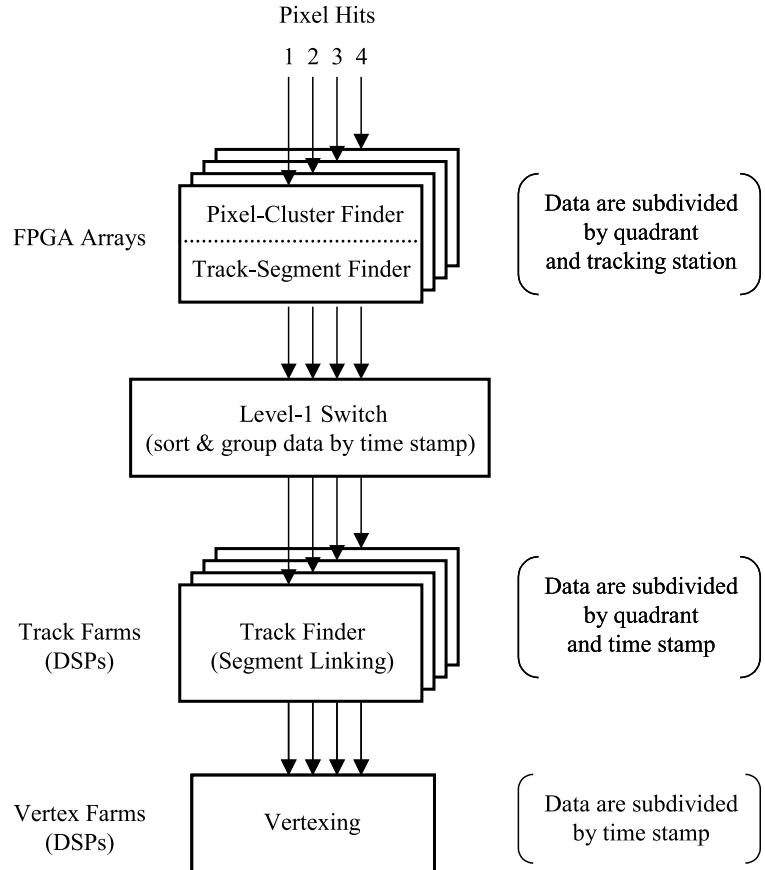
Figure 9.3: Block diagram of the baseline Level 1 trigger.

9.3.2.1 Segment-finding algorithm

As described in Chapter 3, the baseline vertex detector consists of 31 stations each containing two pixel planes, with pixel size $50 \times 400 \, \mu\text{m}^2$. The pixel planes are arranged in doublets, within which the pixel narrow dimension of one plane measures the "bend" (y) view, and that of the other plane measures the "nonbend" (x) view. Track segments are first found in the precision-y view and then confirmed in the precision-x view. For each set of three adjacent stations ($N - 1, N, N + 1$), the steps are as follows:

1. For each precision-y "seed" pixel cluster in station $N-1$, form doublets using precision-y pixel clusters in station N.

2. Form triplets by requiring a confirming precision-y pixel cluster in station $N + 1$.

3. Require at least two confirming precision-x clusters in stations $N - 1, N, N + 1$.

Since magnetic bending is small over these short distances ($\Delta z \leq 8.5 \, \text{cm}$), linear extrapolation is used in projecting from one plane to the next, which is straightforward in FPGAs

122
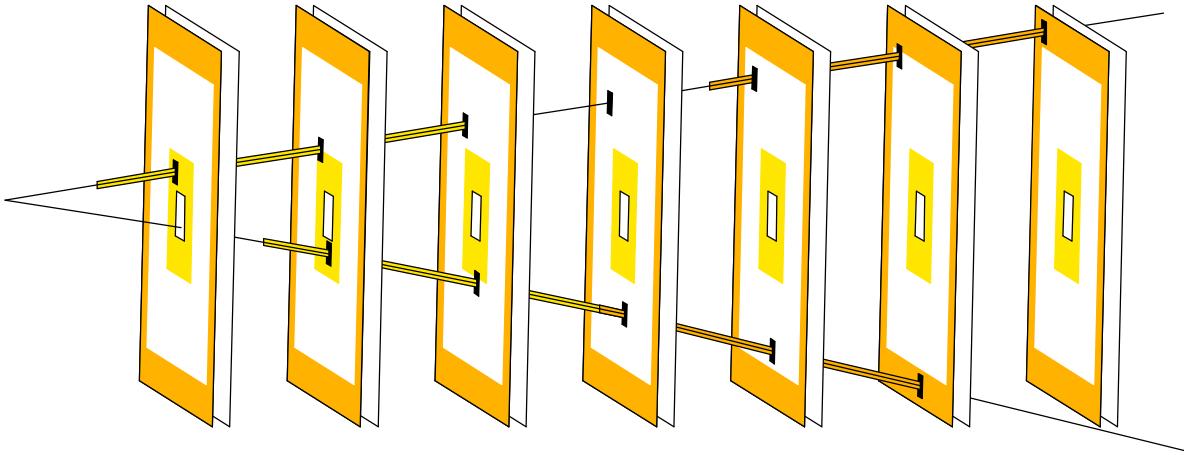
Figure 9.4: Sketch of pixel planes showing (shaded) "inner" and "outer" frames within which track seeds are sought. In the figure, track seeds for the track near the top of the figure are found in the first and last tracking stations. For the lower track the seeds are found in the second and second-to-last stations.

using fixed-point arithmetic. As discussed below, to limit the combinatorics, only "inner" and "outer" pixel clusters are used as seeds. In our baseline studies the inner pixels are taken to be those within 1 cm of the outer edge of the beam hole, and the outer pixels those with $|x|$ or $|y| > 4$ cm (Fig. 9.4).

9.3.2.2 Coping with combinatorics

Since in principle every pixel-cluster doublet could be part of a track, at ≈ 15 clusters/plane it could potentially be necessary to test a few hundred such hypotheses per station per beam crossing. Sub-event parallelism is achieved by subdividing each pixel plane into quadrants and by looking for track segments simultaneously in all four quadrants of all 31 stations in parallel. This 124-fold parallelism enables the segment finding to keep up with the crossing rate.

With ≈ 3.5 pixel clusters per quadrant at 2 interactions per crossing, the number of hypotheses to be examined per quadrant is still ≈ 12, implying a large and expensive installation if all track segments are to be found within one crossing period (i.e. , every 132 ns). To cope with this, our baseline "BB33" algorithm is restricted to finding tracks near their production point, where they should project back into the beam hole, and near their exit point from the vertex detector, where they should project beyond the detector fiducial volume. This means that only those pixel clusters within a frame of "inner" pixels (i.e. , pixels within a specified distance of the inner edge of each pixel plane) and a frame of "outer" pixels (within a specified distance from the outer edge) need to be used as "seeds" for track finding (see Fig. 9.4). These geometric restrictions substantially reduce the number of hypotheses that must be examined and the amount of hardware needed. A further rationale for this choice is that the pixel clusters near the production point give the best resolution for

pixel-plane hit data

plane: N–1   N

inner-pixel cut

projection → AM

Key:
——• write port
——▶ read port

to next stage
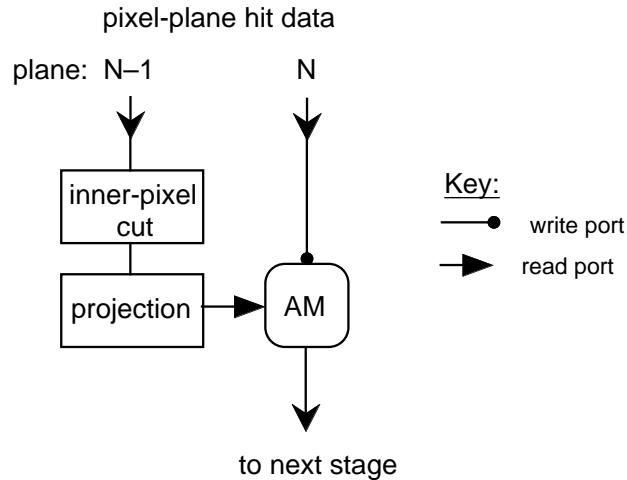
Figure 9.5: Segment-finding "kernel" consisting of an associative memory (AM) with cut and projection operators.

projection to the vertex, while combining them with the clusters near the track exit gives good momentum resolution (needed for an accurate extrapolation to the vertex and proper weighting in vertex fits). Also, the pixel clusters near the track exit provide the best link to the downstream tracking detectors at Level 2.

9.3.2.3   Segment finding with the FPGA tracker

Segment finding is performed for all stations in parallel in "station-quadrant" processors using associative memories [3]. The associative memories are implemented in large FPGAs, for example, the Xilinx Virtex XCV400E, with a cycle time of 20 ns. Details of the implementation are discussed in Appendix A [4]. The key point is that associative memories "unroll" a sequential search algorithm by testing an input value from a "query" list against all the elements of a stored list in parallel. They can thus perform in n cycles a search operation that in typical FORTRAN track-finding code would require a nested DO loop and take $n \times m$ cycles (for example, to search for matches between a list of length n and another list of length m).

To show how it is possible to carry out segment finding with simple hardware in the time available, we must discuss the segment-finding algorithm of section 9.3.2.1 in somewhat more detail. Step 1 of the segment-finding algorithm is carried out as follows by the simple data-processing structure shown in Fig. 9.5:

a) The pixel clusters from plane N are stored in an associative memory.

b) The pixel clusters from plane $N - 1$ pass through a cut operator that compares the position of each cluster to preset limits. Clusters that are in the "inner-pixel" region (Fig. 9.4) are passed to the projection operator.

124

plane: N–2   N–1   N

region of interest

beam axis
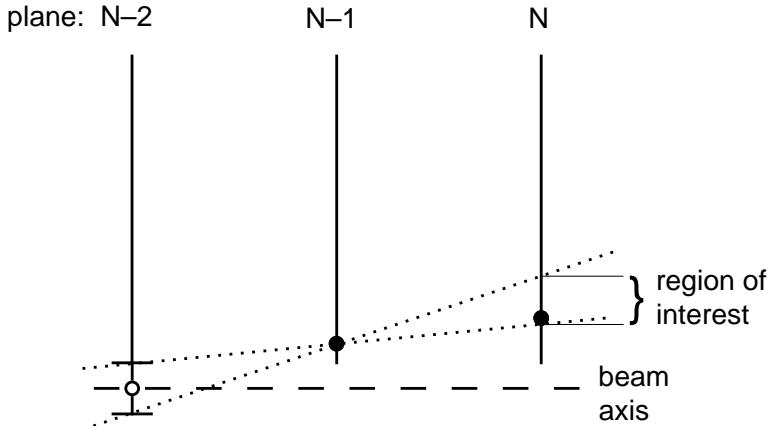
Figure 9.6: Use of a "virtual hit" in finding track segments. The open point with error bars is the virtual hit, while the full points are pixel clusters in planes N and $N - 1$.

c) The projection operator projects each inner pixel cluster from plane $N - 1$ to plane N by using a "virtual hit" that is defined to be on the beam axis at a location that corresponds to the position of plane $N - 2$ (see Fig. 9.6). The goal is to find track segments that are in the beam hole at plane $N - 2$. By projecting a straight line from this virtual hit through a given inner pixel cluster in plane $N - 1$, we select a small region of interest in plane N. Valid track segments must have a confirming pixel cluster within this region of interest. The location of the region of interest is passed to the read port of the associative memory.

d) The associative memory tests whether there is a confirming pixel cluster in the region of interest in plane N. If a confirming cluster is found, its position together with enough additional information to specify the track segment is passed on to the next stage of the segment finding.

We have just described the first step for finding "interior" track segments, *i.e.* those at the track's point of entry into the vertex detector. As shown in Fig. 9.7, the segment-finding "kernel" just described is replicated to carry out the remaining steps of the algorithm. To find "exterior" segments (at the point of track exit), the same operations just described are carried out using outer pixel clusters in station $N - 1$ and a virtual hit located beyond the outer edge of the pixel plane at station $N - 2$. The hardware of Fig. 9.7 performs the complete segment-finding algorithm for both interior and exterior segments.

The time required to process the data from a station-quadrant is equal to the time required to fill the hit lists, at 1 cycle per pixel cluster, plus the time required to query them. Since there is double buffering in the associative memories, the filling time can overlap the processing time for the previous event. The average number of FPGA cycles per event can then be estimated as the average number of pixel clusters per station-quadrant, which in our

BB33 Associative implementation for 1 station
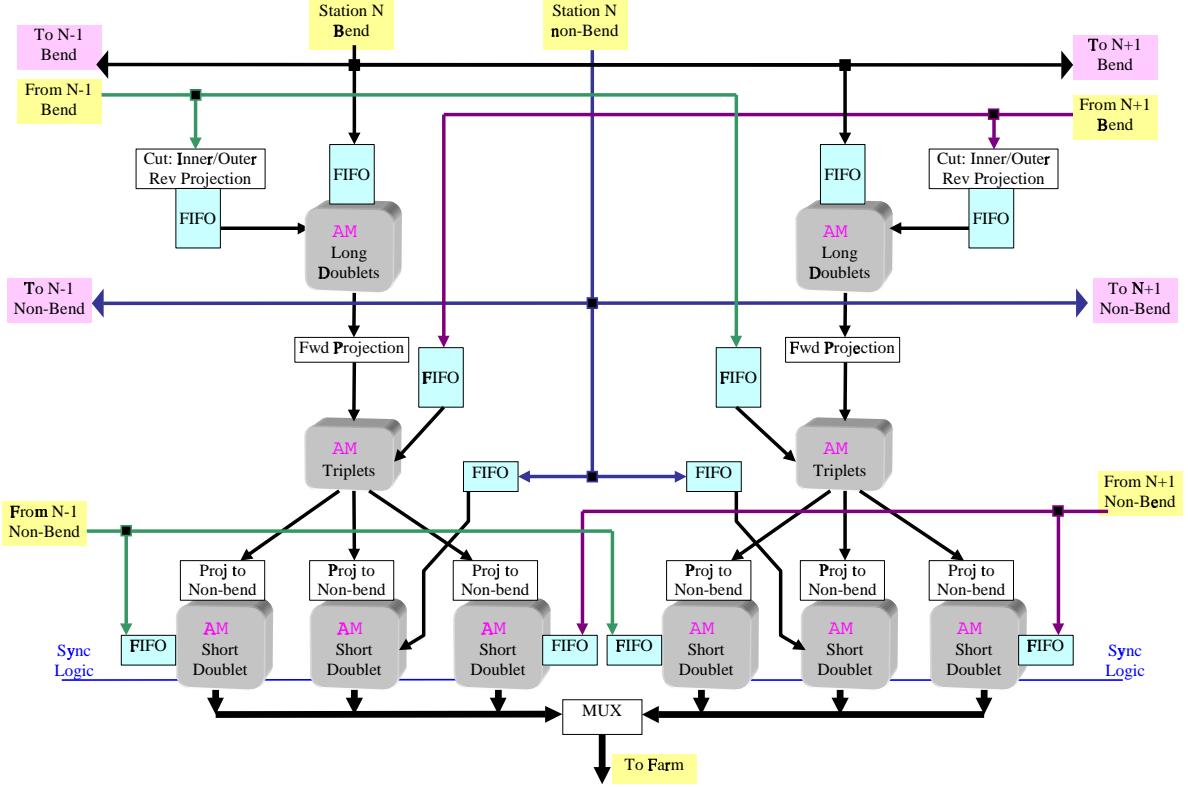
Figure 9.7: Schematic of a "unit cell" of the segment-finding hardware. Note that the drawing is left-right symmetric: the five associative memories in the left half of the drawing are configured to find interior track segments with polar angle $\theta < 90°$ (*i.e.*, segments directed at one arm of the BTeV spectrometer) and exterior segments with $\theta > 90°$, and those in the right half of the drawing find interior track segments with $\theta > 90°$ (directed at the other spectrometer arm) and exterior segments with $\theta < 90°$.

Geant simulation of minimum-bias interactions is 3.5 at an average of 2 interactions/crossing. The average time per event is then 70 ns, well within the 132 ns available.

Note that the operators that perform these operations form a parallel-pipelined dataflow computational structure [5]. Once the pipelines are full, all operators are simultaneously busy, each working on a different piece of the calculation. To smooth out stochastic variations in the dataflow in each pipeline, the operators are provided with first-in/first-out (FIFO) input buffers (see Fig. 9.7). Each of the components in the structure is implemented using a small fraction of an FPGA, so that 16 FPGAs per station suffice to carry out the calculation at the crossing rate without falling behind.

9.3.2.4 Track and vertex farms

The track segments for each quadrant are brought together into track farms, comprised of TI TMS320C67X floating-point DSPs[1] in our design study. In the track farms, the segments within each quadrant are linked together into tracks, and all tracks from all quadrants are then brought together and sent to a vertex farm to be associated into primary vertices. To reduce the complexity of the switch needed for the Level 1 trigger, events are parceled out to eight sets of farms on a round-robin basis according to the beam-crossing time stamp (see Fig. 9.3). There are thus a total of 32 track farms (8 per quadrant) and 8 vertex farms.

The numbers of DSPs required in the track and vertex farms are estimated based on timing studies of the segment-linking and vertex-finding code. Timing analyses show that the optimizing C compiler for the TMS320C67X is not effective at generating efficient machine code for our trigger calculations. The reason is that the trigger code performs numerous sequential and scalar calculations (such as divisions and operations with transcendental functions). We find that for the complicated architecture of the DSP, with its eight pipelined processing units, expert-optimized assembly-language code is considerably more effective, and we have studied code optimizations with the goal of keeping all processing units busy for as large a fraction of the time as possible. This often requires interleaving calculations and rearranging the order of calculations compared to that in the C code. By carefully tailoring the calculations to the specific architecture of the TMS320C67X we have achieved speedup factors of 20 to 50, which leads to a significant reduction in the total number of DSPs required for Level 1. The resulting timings, and the number of processors required to carry out the Level 1 trigger calculations in an average of 132 ns/crossing, are detailed in Appendix A and summarized in Table 9.1.

Table 9.1: DSP processing times and numbers of DSPs required for the track- and vertex-processing farms. The numbers of DSPs are based on the time required to complete the track and vertex calculations in 132 ns, on average.

Task	estimated time/event/DSP	# DSPs
Track processing	55 μs/quadrant	1,668
Vertex processing	66 μs	500
Total DSP count		2,168

9.3.2.5 Performance of the Level 1 algorithm

The studies detailed in the trigger chapter of the simulation section (all carried out at a luminosity corresponding to an average of 2 interactions per crossing) show that the Level 1 algorithm described above finds about 95% of tracks with momenta above 5 GeV/c (Fig. 9.8). (Below this momentum the efficiency drops due to neglect of track curvature in the segment

---

[1]Rated at 1200 MFLOPS with 200 MHz clock.
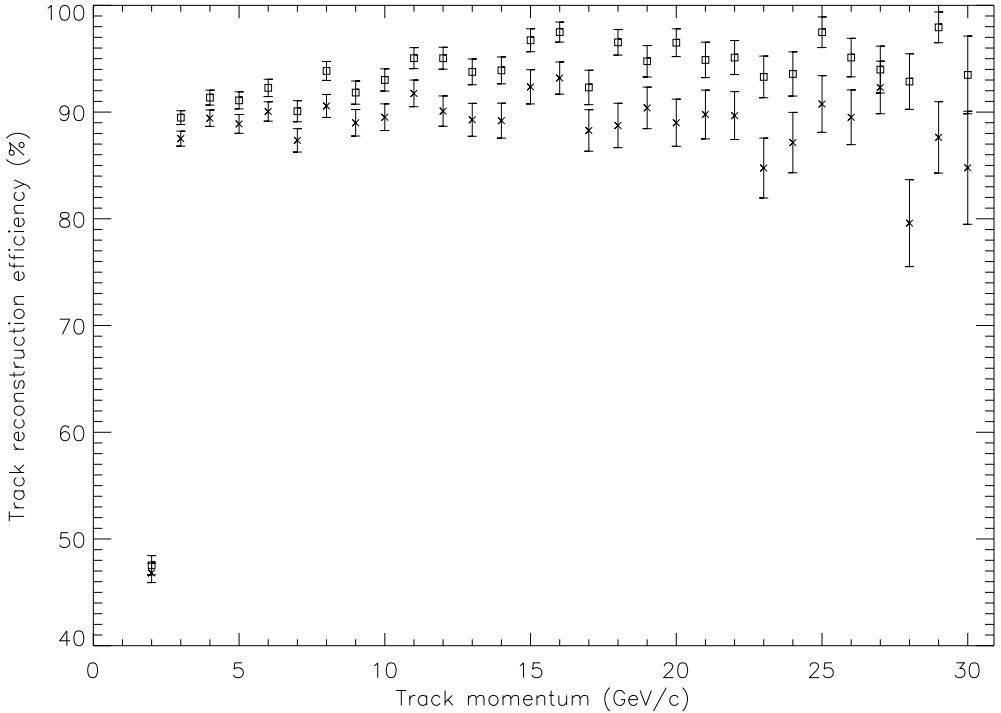
Figure 9.8: Track reconstruction efficiency *vs.* momentum for tracks that are contained within the fiducial volume of the vertex detector. Tracks are required to enter the detector from the beam region, traverse at least four tracking stations, and exit the pixel detector. Data points with squares show the tracking efficiency, which is about 95% above $5\,\mathrm{GeV}/c$, prior to clean-up cuts. Data points with asterisks show the efficiency after clean-up cuts.

finding.) After finding the tracks, we impose simple (and draconian) cuts to remove all tracks that share pixel clusters with any other track. (We may decide to revise these cuts in the future to implement an arbitration scheme that increases the Level 1 track-reconstruction efficiency.)

The Level 1 trigger decision is based on a requirement that at least n tracks (all directed at one arm of the BTeV spectrometer) miss a primary vertex by at least $m\sigma$. This requirement imposes a minimum miss distance for the tracks. However, the miss distance is also required to be less than $2\,\mathrm{mm}$ to exclude tracks that may be associated with other primary vertices. Table 9.2 shows the Level 1 efficiency for various processes of interest requiring at least 2 tracks to miss a primary vertex by 6σ, and Fig. 9.9 shows how efficiencies vary with n and m. In addition to detachment requirements, to improve the point-back resolution [7], displaced-vertex track candidates are required to exceed a transverse-momentum threshold of $0.5\,\mathrm{GeV}$. Efficiencies for typical beauty decays are $\gtrsim 50\%$ while 99% of light-quark events are rejected. Given the order-of-magnitude-higher charm cross section and the larger branching ratios, BTeV's large samples of beauty decays will also be accompanied by comparable samples
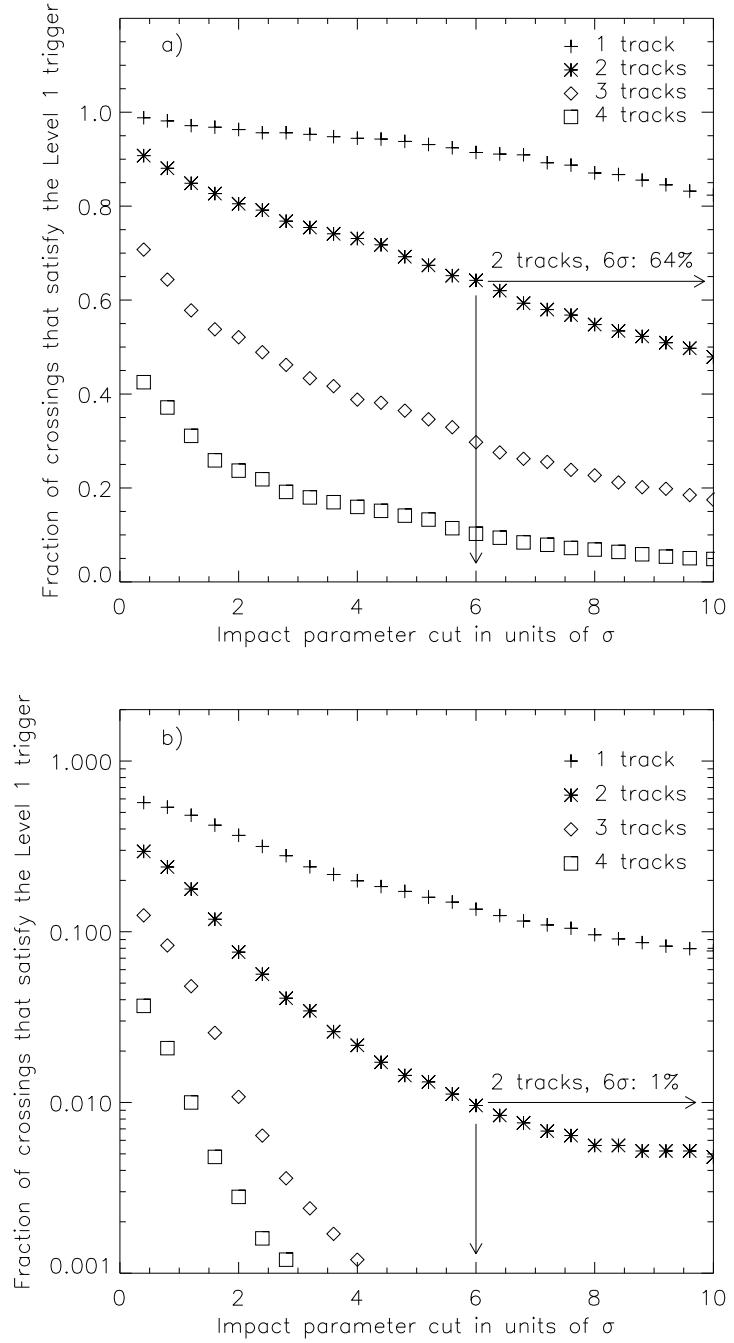
Figure 9.9: a) trigger efficiencies for $B^o \rightarrow D^{*+}\rho^-$, $D^{*+} \rightarrow \pi^+D^o$, $D^o \rightarrow K^-\pi^+$ for various detachment requirements for an average of two interactions/crossing, and b) the trigger response for minimum-bias crossings with the same detachment requirements. Trigger efficiencies are plotted *vs.* normalized impact parameter m for $n = 1, 2, 3$, and 4 detached tracks. The arrows indicate the Level 1 cut requiring $n = 2$ and $m > 6$, which gives a minimum-bias rejection of 99%.

Table 9.2: Level 1 trigger efficiencies for minimum-bias events and various processes of interest that are required to pass off-line analysis cuts. All trigger efficiencies are determined for the Level 1 vertex trigger for beam crossings with an average of two interactions per crossing using the Monte Carlo code shown in the table.

Process	Eff. (%)	Monte Carlo
Minimum bias	1	BTeVGeant
$B_s \to D_s^+ K^-$	74	BTeVGeant
$B^0 \to D^{*+} \rho^-$	64	BTeVGeant
$B^0 \to \rho^0 \pi^0$	56	BTeVGeant
$B^0 \to J/\psi K_s$	50	BTeVGeant
$B_s \to J/\psi K^{*o}$	68	MCFast
$B^- \to D^0 K^-$	70	MCFast
$B^- \to K_s \pi^-$	27	MCFast
$B^0 \to$ 2-body modes $(\pi^+\pi^-, K^+\pi^-, K^+K^-)$	63	MCFast

of decays from directly-produced charm, giving access to possible new physics via charm mixing, CP violation, and rare decays [8].

9.3.2.6 Pixel processors

The pixel processors are the interface between the pixel-detector readout electronics, which provides lists of pixel hits with pulse heights and time stamps, and the Level 1 vertex-trigger hardware. Tracks that pass through the pixel detector typically produce a measurable pulse height distributed over a few adjacent pixels. The measurements in these pixels are combined into pixel clusters that provide the position measurement. These pixel clusters are used for pattern recognition by the Level 1 vertex trigger. The pixel processors also gather all of the data for a given beam crossing for input to the vertex trigger, since the pixel hits do not necessarily emerge from the readout electronics in time order. The pixel processors are designed to be part of the station-quadrant processor boards, each with a total of six large FPGAs. Two of the FPGAs are the pixel processors (one for each pixel plane in a tracking station), while the remaining four FPGAs are used for the track-segment finding. The pixel processors are designed to process hits at a 50 MHz maximum rate to keep up with the interaction rate of 15.2 MHz (beam crossing rate of 7.6 MHz).

9.3.2.7 Global Level 1 trigger

The picture of the Level 1 trigger that we have presented focuses on the Level 1 vertex trigger, which is based on the pixel detector. The actual Level 1 trigger will be more complex, since it includes several different triggers. This complexity is managed by the Global Level 1 Trigger, GLV1.

First, there will be several different vertex-related triggers. We have described a final selection that requires a minimum number of tracks to miss the primary vertex by a given number of standard deviations.. In fact, there will be a variety of such triggers, some accepting events with a few tracks with large detachments and others accepting events with more tracks that have smaller detachments. We also want to record a sample of events that would otherwise have failed the trigger requirements in various ways in order to understand how the efficiency of the trigger "turns on." This last group of triggers may be prescaled.

Second, there will be triggers that involve other detectors. We have previously mentioned the dimuon trigger. In addition to improving the overall efficiency for triggering on states containing J/ψ's, this trigger will permit an independent cross-check on the efficiency of the vertex triggers.

Third, there will probably be triggers that combine information from more than one detector at GLV1. For example, a single-high-p_t-muon trigger would be interesting but may have too high a rate. However, the global Level 1 trigger could accept a high-p_t single-muon trigger if the event also satisfies a relaxed vertex requirement — perhaps one track with a large impact parameter. (At Level 1, one would not know that the high-p_t muon corresponded to the single high-impact-parameter track).

Fourth, there will be special triggers. One example will be a variety of minimum- and low-bias triggers that will be heavily prescaled. Another example will be special alignment and calibration triggers. We will certainly collect special triggers or use minimum-bias triggers to do the quasi-real-time alignment of the pixel detector.

The main physics trigger will not be prescaled. Global Level 1 must have the ability to prescale other less important physics triggers and calibration triggers. We also need the ability at Level 1 to adjust the prescale factors dynamically. For example, we want to be able to reduce the prescale factors on some of the triggers as the luminosity falls. We also want the ability to increase the number of alignment triggers taken at the beginning of a store and then reduce them once enough events have been collected to establish initial alignment constants for the store.

The operation of the Global Level 1 trigger is as follows. The GLV1 receives "trigger packets" from each trigger processor. Each packet contains a header which has "trigger primitives" with a format known to the GLV1. These packets arrive asynchronously. The GLV1 buffers these packets until it receives all the packets it is supposed to receive for a crossing, or until a timeout occurs for that crossing. In normal operation, as soon as all packets are received the GLV1 inspects the headers and generates all the various triggers from truth tables that have been downloaded. It then applies the appropriate prescale to each trigger and takes the OR of the result. If any of the triggers is satisfied, it issues a Level 1 accept. This results in the event data being transferred to the so-called "switch buffer" for transfer to the Level 2 trigger. If no trigger is satisfied, the GLV1 issues a Level 1 reject. This results in the Level 1 event buffers being freed to be used for other events. In general, the trigger processors, which are exchanging messages with GLV1, will have timeouts that are less than the GLV1 timeout. If the allowed time for the arrival of a Level 1 trigger primitive expires before a trigger packet from a processor arrives, an error flag will be generated. A

prescaled sample of these events can then be recorded for further analysis and diagnostics. The rest of the GLV1 decision proceeds as normal.

The fact that the GLV1 receives data packets from the trigger processors allows the processors to send additional data with the trigger primitive information. This could include diagnostic information. It could even include "event information." It is quite likely that we will want to write some of the results of the Level 1 trigger calculations out as part of the event going to the Level 2 trigger. The data may be used for diagnostic purposes or may even be used to "prompt" the Level 2 trigger, for example by directing it to process particular tracks. There are several ways within this architecture to write such data to Level 2. Sending it as part of the trigger-processor packet to GLV1 is only one method.

Since it inspects every crossing, the Level 1 trigger system will be the primary luminosity monitor for the experiment. The GLV1 will keep luminosity statistics and send them to various data logging and monitoring systems.

To minimize the number of designs that must be developed and maintained, the Global Level 1 Trigger will be implemented as a DSP "farm box" identical to the ones used for track and vertex processing.

9.4   Levels 2/3

The Level 2 algorithm refines the tracks found at Level 1 by adding pixel clusters from the planes located between the "inner" and "outer" segments of each track, and (optionally) by adding hits from the first three stations of the forward tracking system (straw tubes and silicon strips). It then performs a Kalman-filter track fit. This improves the momentum resolution from $\approx 6\%$ to $\approx 3\%$, and the vertex-fit Gaussian core to $\sigma \approx 185\,\mu$m. With a requirement on the presence of a secondary vertex or detached tracks having a minimum detached P_t, the result is a joint light-quark rejection of $\approx$500-to-1 per crossing for Level 1 and Level 2 combined, with $\approx 50\%$ overall efficiency for most B decays of interest. The execution time extrapolates to about 15 milliseconds per event on the processor described below. This performance is sufficient for BTeV operation at design luminosity. The Level 2 trigger continues to be studied and refined, both to improve its efficiency and rejection and to decrease its execution time.

The Level 2/3 trigger is implemented as a farm of commercial processors. These could, for example, be INTEL or ALPHA processors running the LINUX operating system. We expect that by the time BTeV runs, processor clock speeds will be at least $2.5\,$GHz. (This is a very conservative assumption, since $800\,$MHz processors are already available in upper-end personal computers.) Given an average decision time of $15\,$ms per Level 2 node, we will need about 2000 of these CPUs.

Assuming existing magnetic-tape technology, the $\approx$10-kHz output rate of Level 2 is about an order of magnitude too high (for media costs) for recording 200-kByte events. Level 3 thus needs to provide a factor $\approx$10 in bandwidth reduction. This can be achieved in various ways, for example by imposing tighter vertex cuts, use of particle-ID information, event topology selection, or event compression (summarizing the raw data). The scenario of

Fig. 9.1 assumes factors of about 4 from the trigger selection and a factor of 4 from event compression (200−300 kbytes/event to 50−75 kbytes/event).

When the Level 2 trigger is satisfied, a Level 2 accept is issued and all remaining data are transferred to the CPU. The buffer memories are freed at this time and the data for that crossing reside only in that CPU's memory. The Level 3 trigger has access to the results of the Level 2 calculations and has an average of about 100−200 ms/event to make its decision. If the Level 3 trigger decides that the crossing should be recorded, it compresses and reformats the event and sends it to a small "data acquisition" farm for eventual archiving to a permanent medium for offline analysis. The output rate to tape is 4000 events/s of an average size of 50 kbytes, which gives a total output rate of 200 Mbytes/s. This can be handled by a farm of about 20−40 processors, with staging disks and two high-performance tape drives attached to each processor.

Bibliography

[1] For example, HERA-B and LHC-b; see talks by E. Gerndt and O. Schneider, Beauty '99 Conference, Bled, Slovenia.

[2] S. Cittolin *et al.*, in *Workshop on Recent Developments in High Energy Physics*, NCSR Demokritos, 9–11 April, 1998, p. 87.

[3] Associative memories (data-storage structures designed for rapid parallel search operations) are finding increasing use in industry; see for example the "content-addressable" memories made by MOSAID Technologies, http://www.mosaid.com/networking/index.html.

[4] See also D. Husby, "Systolic Associative Arrays for Track Finding," http://www-ese.fnal.gov/eseproj/trigger/asmem/asmem.pdf, and "Associative Memory for Track Finding," http://www-ese.fnal.gov/eseproj/trigger/asmem/.

[5] See for example M. H. Schub *et al.*, Nucl. Instr. Meth. **A376**, 49 (1996); W. Sippach, G. Benenson, and B. Knapp, IEEE Trans. Nucl. Sci. **NS-27**, 578 (1980); R. G. Cooper, IEEE Trans. Comp. **C-26**, 1123 (1977).

[6] D. Husby *et al.*, *Nucl. Instrum. Meth.* **A383** (1996) 193.

[7] W. Selove, in *Proceedings of the Workshop on B Physics at Hadron Accelerators*, P. McBride and C. S. Mishra, *eds.*, Fermilab-CONF-93/267 (1993), p. 617.

[8] D. M. Kaplan and V. Papavassiliou, in **CP Violation**, X.-H. Guo, M. Sevior, and A. W.Thomas, eds. (World Scientific, Singapore, 2000), p. 116; D. M. Kaplan, in *Proc. Symposium on Flavor Changing Neutral Currents: Present and Future Studies*, Santa Monica, CA, 19–21 Feb 1997, D. B. Cline, ed. (World Scientific, Singapore, 1997), p. 81.

Chapter 10

Data Acquisition System

In BTeV, the Data Acquisition System and the trigger are intimately connected. In the preceding chapter, we focused on trigger requirements and algorithms. In this chapter, we describe the BTeV Data Acquisition System - the DAQ, concentrating on the data movement from the detector, through the various stages of the trigger, and finally onto mass storage.

10.1   Data Acquisition

The BTeV system will digitize, sparsify, and transmit data at the beam crossing rate of 7.6 MHz into off-detector buffer memories. This approach is taken because a large subset of the data is used in the first level trigger, the data must be digitized in any case, and a very sophisticated first level trigger is planned. With this approach applied to all subdetectors, the first level buffers can hold many more beam crossings of data than those found in typical front-end integrated circuits designed to have (typically small) on-chip buffering. This means that the decision time of the Level 1 trigger can be extended by as much as two orders of magnitude, if desired, allowing for much more sophisticated trigger processing than would be possible given a short (typically a few μs) Level 1 latency.

Three distinct logical trigger levels are described in section 9.2. From an engineering perspective, there is little difference between Level 2 and Level 3; these levels are both executed in general purpose processors. The only distinction between these levels is that the start of the Level 3 processing presupposes that ALL data have been transferred to the processor, whereas the Level 2 processing operates only on a subset of the total data. Thus, triggers in the BTeV system are separated into two physical levels, referred as L1 and L2/3. This is typical of most proposed large scale data acquisition systems and acknowledges the greatly increased performance and decreased cost of general purpose processors. In effect, much of the early and intermediate processing in previous trigger architectures has been moved into the first level hardware, Level 1 in our case, while the remaining processing occurs in a greatly expanded array of general purpose processors.

Following the Level 1 trigger, the remainder of the system is sized for a minimum L1 rejection of 75:1, which represents an L1 accept rate of $\approx$ 100 kHz. This rate must be

directed to the L2/3 processors though a switch. Moreover, the L2/L3 processors must be able to make decisions at this rate or data will be lost. Studies based on a preliminary Level 2 trigger algorithm and an estimate of the Level 3 processing time based on experience with our own offline analysis and from other experiments indicate that this can be achieved with ≈ 2500 general purpose processors of about 2500 MIPS each in the L2/3 processor array. The use of general purpose processors does not preclude the use of dedicated L2 trigger logic in the form of attached coprocessors.

The specifications shown in Table 10.1 are used as the baseline for the BTeV data acquisition. These numbers represent both arms of the detector.

event size	100 kBytes
number of detector data links	5000
number of L1 data buffers	400
number of L2/3 data links	64
number of L2/3 processors	2500

Table 10.1: Estimates of Hardware for BTeV Trigger and Data Acquisition System

The total system buffer memory, assuming 400 L1/switch input buffers and 64 switch output buffers, is almost 100 Gigabytes. Buffers in the L1/2/3 processors will push total system memory to approximately 400 Gigabytes. A block diagram of the trigger and data acquisition system is shown in Fig. 9.1.

10.1.1 Detector Multiplexing

To efficiently balance data rates on the detector data links, a relatively large number of detector channels will be multiplexed into each high speed link. In the case of the RICH subsystem, for example, it is expected that approximately 1200 single bit signals will be multiplexed and encoded into one link. The input to the modules driving these links will depend in detail on the detector subsystem, but the use of programmable logic will allow for some variation in signal widths and sparsification algorithms while using common hardware. The detector multiplexer drives the detector data link and receives the coded clock signal for fanout to the front-end components.

10.1.2 Detector Data Links

Each detector data link must transfer approximately 100 MBytes/sec. This rate can be accommodated by a single 1-2 Gbps fiber. The choice of technology and link speed depends on the physical placement of the first level buffers. The total bandwidth of the detector data links is about 500 GBytes/sec.

10.1.3   L1 and Switch Input Buffers

Each Level 1 Buffer must accept data from a subsection of the detector at a 1200 MByte/sec average rate and the data must be held in this buffer pending a Level 1 trigger decision. The BTeV L1 trigger is based in large part on track reconstruction in the vertex detector and decision times may vary over a wide range depending on event complexity. For this reason, it is desirable that the L1 buffer controller support receipt of both asynchronous and out-of-order L1 accepts and rejects. Following each L1 accept, the data are transferred to a switch input buffer. The function of the switch input buffer is to hold data until an L2/3 processor assignment is made, the data have been requested by the processor, and the switch rotation allows transmission of the data. Data requests to the switch input buffer also arrive asynchronously and in arbitrary order.

As a practical matter, the L1 buffer and switch input buffer may use the same physical memory, with the data "transferred" between buffers by reassignment of pointers. The size of an individual data buffer is expected to be approximately 128 MBytes.

Data from some detector data links will also be used in the L1 trigger. In this case, the L1 buffers receive data from the L1 trigger instead of the detector.

10.1.4   Ring Buffer Interconnects

Our design approach is sufficiently flexible to accommodate different options for data transmission to the L2/3 processors. If all data is transmitted to the L2/3 processors following an L1 accept, the most natural arrangement of L1 buffers is to place an equal number on each switch input link. If the data is transmitted in steps (staged readout), placement of the L1 buffers should be optimized for the frequency of readout. Using a ring architecture achieves the same multiplexing functionality as a fixed backplane bus, while also allowing reconfiguration of the L1 buffers to match changes in trigger rate, switch size, and access frequency of the buffers.

The ring and switch interconnects utilize 1 Gbps serial links based on the same physical layer standard as Gigabit Ethernet, Fiber Channel, and "Infiniband". All data are transferred in fixed length packets, similar in size to an ATM cell.

10.1.5   Staged Readout

To reduce the size of the data switch, the Level 2 trigger may operate in stages. In each stage, it would then request only enough information from the switch input buffers to make a decision to continue to the next stage of processing. Events may be then rejected at any stage in the Level 2 analysis, with the remaining data cleared from the input buffers without being transmitted through the data switch.

The reduction in data switch bandwidth requirement is offset by the increase in bandwidth required for the data request messages flowing in the reverse direction. In addition, the input buffer depth must be increased significantly to provide greater latency for event data used in later stages of the L2 algorithm. Buffer management logic also becomes more

complicated. The cost reduction comes from the potential decrease in the number of required output buffers. An analysis of the minimum number of required output buffers and the expected fraction of data needed in the average L2 calculation will be performed to determine if the staged readout approach is cost-effective.

10.1.6 Data Switch

A simple, unidirectional, packet synchronous, TDM switch has been shown to be the most efficient for event building applications. The switch configuration follows a standard "barrel shift" rotation and can be expanded to any size in NlogN, where N is approximately the total data rate required by the Level 2/3 trigger divided by 100 Mbytes/sec (switch maximum throughput) times a safety factor. We expect to eliminate 80-90% of the events at level 2 based on inspecting a very small amount of the data from each event. Under rather conservative assumptions, the BTeV data switch is expected to require a maximum of 64 channels, which can be implemented using two stages of 16 X 16 switch modules.

The advantage of this switch is that it requires no complex internal control. Source and destination addresses contained in the data packets are used only to specify buffers on a common ring and are not needed by the switch for data routing. The switch is unidirectional, but packets are recirculated by connecting the last output buffer to the first input buffer in each ring. This provides a mechanism for input-to-input message transfers (throttle requests to the global trigger), output-to-output message transfers (L3 accepted events to logging processors) and output-to-input messages (event requests to the global trigger/event supervisor).

Because the switch is easily expanded, full implementation can be delayed until the accept rates of the L1 and L2/3 triggers are determined, and the switch sized accordingly. The cost of the switch is not a large factor in the overall system cost.

10.1.7 Control Network

A broadcast network is needed to distribute the accept/reject messages to the switch input buffers. This network is implemented as a second ring which follows the same physical path as the ring buffer interconnects.

10.1.8 Switch Output Buffers

The switch output buffers are similar in function to the switch input buffers. This is the point where all the requested data from a single event is received from the switch and forwarded to a processor. Connection to processors is by way of Fast Ethernet links, with 64 links per switch output buffer. Up to 4096 processors can be connected to 64 switch output buffers, and additional buffers can be added to each ring if the processor array is expanded.

10.1.9 Data Logging

Events accepted by the Level 3 trigger are sent to data logging processors. These processors are attached to switch output buffers in the same way as any other L2/3 processor. Accepted events are placed back on the ring and recirculate through the data switch to the appropriate output port. Given the high trigger rejection, this utilizes approximately 1% of the data switch bandwidth while avoiding the need for a separate data logging network.

10.1.10 Detector Control Links

All detector subsystems transmit data at the beam crossing rate. This simplifies the delivery of clock and control signals to the detector front-end components because L1 and L2 accept/reject signals are not required at this level. It is assumed that a 53 MHz coded clock will be sufficient to provide crossing and synchronization information. Clock and data recovery should be provided by the detector multiplexing module and then delivered independently to the front-end.

10.1.11 Diagnostics

All system buffers have a control link for the purpose of sending or receiving data request messages. This provides a convenient path for system diagnostics at little additional cost. We expect to make widespread use of embedded processing and embedded servers in these modules to allow remote test and status monitoring. An additional low speed interface (USB) is available for stand-alone diagnostics.

10.1.12 General Operation

At startup, buffers can be initialized through the control link. The data switch is initialized by filling the rings with empty data packets, after which it operates in a packet synchronous mode. The front-end systems are synchronized through the clock distribution system.

Event fragments are timestamped with the crossing number which is used for all Level 1 trigger accept/reject messages. Events accepted at L1 are tracked through the remainder of the system by the L1 event number. An L1 accept causes the buffer control logic to move a memory pointer from the crossing number list to the L1 event number list. At the same time, the L1 event number is added to the list of events available for L2/3 processing.

An event request from an L2/3 processor is sent to the Event Supervisor (see below). The Event Supervisor returns a packet containing an L1 event number, along with any trigger information considered useful to the processor. At this point the processor is responsible for disposition of the event. It may request some or all of the data from switch input buffers for use in the L2 decision. Following an L2 accept or reject the processor makes a final data request to all remaining switch input buffers. Buffers then delete or transmit/delete the event data. Each switch input buffer receives exactly one data request message per event.

10.1.13 Component Placement

The detector data links and L1 buffers represent a large part of the overall data acquisition system cost. Placement of the buffers influences the technology and cost of the links. The buffers will occupy the equivalent of 25 6U subracks. The L1 subsystem triggers should be located in the same area as the L1 buffers. These components would be placed outside of the collision hall to facilitate maintenance and to allow reconfiguration of rings based on changes in trigger rate and buffer access frequency. This requires the use of optical links. Placement of all other data acquisition components (switch, output buffers and processors) is not critical, since there are a relatively small number of interconnects.

10.2 Event Supervisor and Monitor

As our design has developed, the Global Level 1 Trigger system began to acquire many of the functions of an 'event supervisor and monitoring system'. In this section, we will describe this functionality. We will refer to the hardware that does these tasks as the Event Supervisor and Monitor, ESAM. Whether, in the end, it is provided by the same hardware responsible for the Global Level 1 Trigger or it assumes a separate identity is an implementation detail.

The ESAM will maintain a list of events which have satisfied the Level 1 trigger. When a L2/3 processor becomes available to process a new event, it sends a request for an event to the ESAM. The ESAM then sends a data packet to the processor which contains the event number to analyze. It also includes the status of all the triggers for that event so that the processor can execute different code depending on how the event triggered. It may also send more extensive event data from the trigger processors. Once the ESAM has sent an event number and associated data to a Level 2/3 processor, it is done with that event.

ESAM is the natural location for accumulating statistics on the performance of the Level 1 trigger. It will periodically collect statistics from GLV1 and send this data up to online analysis computers.

One interesting application of this is in luminosity monitoring. Since the BTeV Level 1 trigger inspects every crossing, it will be possible to count the number of crossings with no interaction as well as the number of interactions in each crossing that has at least one. Given the large acceptance of the pixel detector for inelastic collisions, we believe that the Level 1 trigger system will function as a highly effective, real time luminosity monitor. We will set up a separate data path, using fast Ethernet, from ESAM to a small dedicated processor to accumulate statistics from Level 1 luminosity information, to archive the information, and to run various status displays for operations.

ESAM will also play a role in data throttling – that is trying to reduce the rate of data sent to the upper trigger levels if they begin to fall behind and the switch buffer memory approaches its capacity. The 'switch buffer memory' which holds the event data while waiting for a L2/3 processor to become free will be able to issue a 'full warning' to ESAM when it becomes 90% full. ESAM will send this to Global Level 1 which can attempt to undertake some action to alleviate the problem. For example, it could begin to adjust prescales to

reduce the total trigger rate. It could also stop sending events. When the 'switch buffer memory' is completely full, it will issue a 'full' indication. ESAM will send this to the GLV1, which will stop issuing Level 1 accepts until this condition is cleared. This will result in deadtime and ESAM will have to keep track of how much deadtime there is for purposes of normalization and for diagnostic purposes. It is, of course, a BTeV design goal for this to almost never occur.

We are also discussing 'partitioning.' Partitioning refers to the ability to divide the Trigger/DA system into more than one semi-independent piece, each of which approximates a completely independent trigger/DA system. This capability is especially useful in the debugging stage of an experiment. In this stage, some subsystems may not be installed, some may not yet be operational, some may be down for maintenance, and some may be hanging or crashing. It is important to be able to work on one or a collection of subsystems independently of the state of the others. The BTeV architecture permits a reasonable level of partitioning. The ability of Level 2 processors to request subevents rather than whole events lends itself to partitioning. With the appropriate software, we can extend this capability to Level 3 so a process can proceed when it gets just the subset of data it needs. On the other hand, the requirement that an event be given to one and only one processor creates some difficulties for partitioning.

In our current plan, the ESAM will maintain separate trigger lists for different 'event' types. It will also maintain separate lists of event numbers for each triggering event type. It will have an algorithm for assigning events satisfying triggers in more than one list. The algorithm will probably be based on the desired event rates and priorities for each event type. Level 2 processors will request an event of a specific type. ESAM will return an event number from the list for that type. The Level 2 processor can then request the subevents that it wants to analyze. If the 'switch buffer' is configured so that it waits a certain minimum interval before it sends an event (so that it is sure that it has the whole event), and if it can send a message saying that it doesn't have a requested subevent, then the fact that some (presumably) unneeded subevents are missing does not cause a problem. With this scheme, subgroups can have their own triggers and select the subset of the data they want to read independently of each other. However, since each event will go to only one processor, ESAM will have to decide how events satisfying the requirements of multiple trigger lists should be apportioned.

10.3   Level 2/3 Trigger

As discussed above, the Level 2/3 Trigger is a processor farm. The baseline design of the farm calls for 2000-4000 general purpose processors, such as INTEL/Pentium PCs, running Linux. Each processor runs the same analysis code. Requests for new data are sent to the Global Level 1 Trigger, GLV1, which coordinates the initial data requests for each processor in the Level 2/3 Trigger. The Level 2/3 analysis software for BTeV performs data analysis that is usually part of the "offline analysis" in other experiments. Consequently, the Level 2/3 Trigger can select events using cuts that are usually reserved for offline analyses, and can

transform event data into a condensed format (usually referred to as a DST format) before sending the data to event-logging processors.

An important aspect of the Level 2/3 Trigger architecture is that each processor is capable of making trigger decisions at both Level 2 and Level 3. The distinction between Level 2 and Level 3 depends on the memory buffers in which the data for an event reside, and will be described below. A second aspect of the trigger architecture is that each event is assigned to a single processor in the PC farm, so that a single processor selects or rejects an event at Level 2 and, if it passes Level 2, at Level 3.

The Level 2 analysis for an event begins when an idle processor sends a request to the Event Supervisor for a new event. The ESAM, which maintains a list of events that have satisfied the Level 1 Trigger, responds by sending a data packet to the processor. The data packet contains the event number, and Level 1 trigger information. The processor begins the Level 2 analysis of the event based on the Level 1 trigger information, which summarizes all of the trigger conditions that were satisfied at Level 1. In addition to the trigger information, the processor usually requires data from various detectors to complete the analysis and arrive at a trigger decision. The data reside in switch buffers. The processor requests data from these switch buffers directly. For example, the processor may send a request for data recorded by the two forward tracking stations that are closest to the dipole magnet. The requested data are sent from the switch buffer to the processor's memory and the memory it occupied in the switch buffer is freed. At this point, part of the event is in the processor memory and the remainder is in the switch buffer. In this example, the added data blocks are used to improve the momentum determination of tracks found by the vertex detector, so that the processor can calculate more precise impact parameters to select tracks coming from B decays. After the data are received and processed, the processor may request additional data (such as from the other forward trackers or from the muon detectors, in the case of a Level 1 muon trigger), or complete the Level 2 analysis by sending a trigger decision to the switch buffer.

A processor that has completed the Level 2 analysis for an event sends one of two trigger decisions to the switch buffer. If the event is rejected, a "L2 reject" is sent and any remaining data blocks for that event are deleted from the switch buffers, freeing that memory for new events. The processor is ready to process a new event. If the event is accepted, a "L2 accept" is sent, and the switch buffer sends the rest of the data data blocks for that event to the processor for subsequent Level 3 analysis, and removes that data from the switch buffers. This marks the beginning of the Level 3 trigger. At this stage, the data for an event reside entirely in the memory of one processor (the same processor that performed the Level 2 analysis for the event), and nowhere else.

A processor that performs the Level 3 analysis for an event can perform a complete analysis using all of the data that are associated with that event. The data analysis is comparable to "offline" analyses in other experiments, and includes track reconstruction, vertex reconstruction, muon identification, and (possibly) charged hadron identification. Events that are rejected by a processor at Level 3 are simply terminated by that processor. Events that are accepted are sent to data-logging processors.

10.4 Data Logging and Online Event Analysis

Each event is handled by only one Level 2/3 processor. (i.e., an event is not scattered across processors nor is it copied to more than one processor). All accepted events need to be moved out of the L2/3 processors and onto tape for long term data storage, and a fraction of the events need to be made available for online monitoring (by "consumer processes").

Events coming out of the Level 3 processors will not be raw data, but already processed data (except for a highly prescaled sample used for monitoring the data reduction algorithm itself). This should shrink the event size from 200 kBytes to 50 kBytes. Assuming an event size of 50 kBytes, and a 4 kHz event rate to the loggers, the data rate to tape is on the order of 200 MBytes/sec. The consumer processes will also add an additional 5-10% of throughput.

Since the data logging rate out of each Level 3 processor is small, it is less cost effective to attach logging media to these individual nodes than to provide a small number of separate logger nodes. The necessary bandwidth to the logger nodes is also small (2-3%) compared to the raw data coming up the Level 2/3 farms, so the same switch could be used to pass the event to the logger nodes as well. Alternatively, a dedicated network could be provided between the Level 3 processors and the logging nodes. In the current plan, a single event is routed to the next free logging node but buffering several events in the Level 3 processors and sending them out together is possible if it turns out to be beneficial.

The number of logging nodes themselves is clearly a function of the data rate. It is assumed that data will first be buffered to disk before being written to tape. This serves a dual purpose; buffering enough data before a transfer to keep the tape drives streaming, and protecting against tape media errors and tape drive failures. With current market technology, a single processor node can handle 10 Mbytes/sec to tape including the initial disk write. We will therefore require a minimum of 20 data logging nodes. We are planning for twice this number to have a safety factor and to be able to handle failures.

Consumer events should not interfere with data logging rates and additionally may contain samples of rejected or flawed events that we may want to look at but not to write out to permanent storage. The consumer processes will therefore run on separate nodes from the logger processes.

10.5 Front End Electronics

In this section, we discuss those aspects of our front end electronics which are common among systems. Detailed information on the readout of each detector can be found in the chapters describing the individual detectors. We conclude with a brief discussion of software support for electronics development.

10.5.1 Infrastructure and Other Support Systems

A goal of the BTeV experiment is to minimize hardware and software development while meeting all the requirements of the experiment. This goal can be accomplished by:

- Choosing cost-effective, commercially-available hardware and software whenever possible;

- Using hardware and software developed elsewhere whenever possible; and

- When hardware and software need to be developed for BTeV, developing them in such a way that they can be used in the maximum number of BTeV front-end, trigger, and data acquisition systems.

Commonality in hardware and/or software when possible in BTeV's front-end, trigger, and data acquisition electronics will save not only substantial amounts of money but valuable personnel resources. A BTeV workshop was held to understand

- what already-developed electronics and software can be used in BTeV and

- where hardware and/or software; and commonality is possible in BTeV's front-end, trigger, and data acquisition systems.

The following subsections describe our choices for some infrastructure items and other support systems.

10.5.1.1   Electronics Packaging - Subracks and Related Infrastructure

Using a mechanical packaging and bus protocol standard that is commonly used internationally offers several advantages over the lifetime of an experiment. Cost and personnel time savings are the two most important advantages. Standards are well-documented and stable. Various electronics and mechanical components supporting the standard are commercially-available from industry. For example, subracks, processor modules, (module size) adapters, extenders, subrack power supplies and interface ICs are commercially available items not requiring development. This leaves us free to concentrate only on the development of specific front-end, trigger, data acquisition, controls and monitoring electronics. By choosing a commonly-used packaging and bus protocol standard, implementers, as well as people who maintain the electronics over the lifetime of the experiment, have a mature development and testing hardware and software environment. If the standard has been used in previous experiments at Fermilab, we can realize considerable cost and personnel time savings because local expertise is already available and infrastructure is in place.

We have chosen the VME64 Extensions (VME64x) subrack mechanical packaging and bus protocol standard as the baseline implementation for the BTeV front-end, trigger and data acquisition electronics. This is an American National Standards Institute (ANSI) standard, ANSI/VITA 1.1-1997. This standard and its physics community VME64xP extension are being used for all of CDF's and a majority of D0's Run II electronics. Over 250 VME64xP subracks have been purchased by CDF and D0. Other standards used in conjunction with this standard are:

1. IEEE1101.10 - the subrack module mechanical standard for VME64x;

2. IEEE1101.11 - the subrack rear I/O board (transition module) mechanical standard for VME64x;

3. ANSI/VITA 1.3-1997 - the VME64x 9U x 400mm module format standard;

4. ANSI/VITA 25 - VISION, a software VME API (subroutine calls) standard; and

5. ANSI/VITA 23 - VME64 Extensions for Physics (VME64xP), a VME64x-compatible standard which extends VME64x by, for example, specifying 7U and 10U subrack implementations and specifying mandatory control and status registers.

These standards offer many advantages for BTeV, which will take data for long periods of time in a challenging environment. Attention to cooling, power, electronic noise rejection and ease of maintenance are all requirements which are addressed by the standards.

An alternative method of packaging BTeV electronics has been proposed. This method draws on products developed in the world-wide, very price-competitive, PC packaging industry. No communications busses would be used. Slow controls for component initialization, downloading and monitoring would be done by interconnecting Ethernet between PC packages. Electronics would thus not be modularized but, for example, each set of front-end electronics would be housed in its own PC package. Power to that package would be via a 110 VAC line cord and power supplies within the package. This option potentially offers cost savings over bussed systems but would require the development of a new infrastructure for hardware testing and software development. This option would only be adopted if it resulted in substantial cost savings over the life of the experiment taking into account all development, support, and maintenance issues.

10.5.1.2 Data and Clock/Timing Links and Link Interfaces

Possible and probable areas of hardware commonality in data and control/timing links and other components include:

- Data links from front-end electronics to Level 1 buffers;

- Data links from the Level 1 buffers to the event building switch;

- Data links from the event building switch to event buffers preceding online processors;

- High-speed front-end data links from on-detector electronics, possibly rad-hard, to local electronics;

- Low-noise data links from on-detector electronics, possibly rad-hard, to local electronics;

- Data links to Level 1 trigger subsystems and from those subsystems to the Global Level 1 trigger;

145

- Control/timing links to front-end electronics and trigger subsystems; and

- Level 1 buffer electronics.

Two ongoing developments and one existing Hewlett Packard (HP) data link series of products should satisfy most if not all of the data and control/timing link applications given above. HP's G-links are transmitters and receivers which operate up to 1.5 Gigabits per second over fiber or copper. CERN is developing a G-link compatible radiation-hard driver for use in its LHC CMS (HCal and ECal) and ATLAS experiments. Fermilab is assisting that development, and we plan to use it for pixel data readout links, and possibly other subsystem data readout links. CDF SVX Run II electronics required the development of a radiation-hard, low-noise transceiver. This transceiver will go into production soon. It is compatible with the IEEE-1596 LVDS (Low-Voltage Differential Signalling) standard developed for the Scalable Coherent Interface standard. This is a differential current standard intended to provide low-noise transmission of digital signals at very high speeds for a few tens of meters maximum.

We will adopt a minimum set of "BTeV data link" standards and, hopefully, a single "BTeV control/timing link" standard. The types of "BTeV links" will include:

- Serial optical G-link 1.3 Gigabit per second data link transmitting up to 20 bits every 1/53MHz (radiation-hard and non radiation-hard); and

- Parallel copper LVDS-like to LVDS 53MHz data link (radiation-hard and non radiation-hard).

The parallel copper LVDS-like to LVDS data link would be used to move data a short distance from the detector to an area near the detector where electronics is more accessible. It provides for relatively low-power, low-noise data transmission off the detector. An alternative for this application is an array of Vertical Cavity Surface Emitting Lasers (VCSELs) and an optical ribbon cable link. VCSELs are inherently radiation hard. If this approach is cost effective and does not add more mass than desired, only minimal electronics need be on or near the detector. All the other electronics (e.g., data multiplexers, Level 1 buffers, trigger electronics, etc.) can be in the counting room and thus very accessible.

Our front-end and trigger subsystem implementers will easily be able to interface their electronics to BTeV standard readout and data acquisition electronics and control/timing signals. This will be accomplished by providing them with data and control/timing link daughterboards. Along with the daughterboards, the implementers will be provided with mechanical and electrical specifications for the daughterboards and component and artwork specifications for circuitry on their boards used to interface to the daughterboards. The following daughterboards are under consideration:

- Serial optical G-link or G-link compatible data link transmitter daughter-board;

- Parallel copper LVDS-like data link transmitter daughter-board; and

- Serial optical G-link data receiver control/timing link daughter-board

10.5.1.3 Slow Controls and Monitoring

In an experiment the size of BTeV, several hundred devices need to be controlled (e.g., high-voltage systems, laser pulsers, interlock systems, etc.). Several thousand 'points' (e.g., power supply voltages, temperatures, gas mixes, interlocks) need to be monitored at regular intervals. Alarm or caution situations need to be established via hard-wiring or in software for equipment and/or personnel safety. Past experiments at Fermilab have chosen to use commercial or in-house-designed hardware and in-house- developed software for slow controls and monitoring systems despite the fact that total process automation solutions (hardware, software and user interfaces) are provided by a multitude of companies worldwide. The beamline cryogenics controls and monitoring system was the first system at Fermilab to use an industry total process automation system. Their hardware controllers and input devices were purchased from Moore Products and compatible software and graphical user interfaces from Intellution. The success of this system was remarkable. Mechanical engineers bought it and implemented it with no Fermilab professional software or electronics engineering help. Significantly fewer people were required to operate the system. As a result CDF will use an identical system for its Run II slow controls and monitoring system. BTeV's baseline slow controls and monitoring system is based on this system.

10.5.1.4 Clock and Timing Distribution

The BTeV experiment is different from all other current and past Fermilab experiments in that a) tracking (pixels) will be used in the lowest level trigger and b) all front-end subsystems will be read out into Level 1 buffers at the 132 nanosecond bunch crossing rate. Most front-end subsystems will require only a bunch crossing clock and a 53 MHz accelerator clock. Some will operate asynchronously not needing a bunch crossing clock. For Run II, CDF has used a modular subsystem for distributing clock and timing information to its front-end and trigger subsystems. Differential ECL signals over high-quality cables are used for long-distance runs from the clock/timing sources to the collision hall. Low-current differential signals (LVDS drivers and receivers) are used to transport clock and timing signals to on-detector front-end electronics and counting room electronics. We will try to use as much of the work done by CDF as practical for its clock and timing distribution subsystem.

An alternative approach to sending clock and timing information to front-end and trigger subsystems is used in the Run II CDF SVX system. A conditioned accelerator clock (with jitter essentially removed) is used to clock timing data (e.g., bunch crossing) into HP G-links via optical cables. Optical splitters can be used to distribute the clock and timing information. At the output of the G-link receiver, the clock is rederived from the output data strobes of the G-link receiver. The bunch crossing timing signal is received as a G-link receiver data bit.

10.5.1.5 Electronics Cooling and Rack Protection

Several systems for cooling and protecting electronics racks exist at Fermilab. Our cooling system will be designed so that the exhaust air from a VME subrack will not exceed 95 degrees Fahrenheit. The fans will minimally provide 400 linear feet per minute of air flow. The air/water heat exchangers will reduce the temperature of 95 degrees Fahrenheit input air to 85 degrees Fahrenheit at the above air flow. There is no reason we can't use either CDF's or D0's Run II electronics cooling and rack protection system. This system will be connected to the slow controls and monitoring system described above.

10.5.2 Software for Module and Subsystem Development and the Online Data Acquisition and Readout System

BTeV front-end and trigger subsystem developers can use existing software for developing modules and subsystems. An infrastructure for testing VME-based boards has been developed for CDF, and is the foundation for testing CDF's Run II SVX, Muon and Calorimetry system hardware. This framework, called CDFVME, is Java-based code that communicates to the test subrack(s) via CORBA, a powerful object request broker, which here is used to provide the functionality of a remote procedure-like software tool. The CORBA implementation is ROBIN, and requires a VME processor in the subrack running the VxWorks operating system. VxWorks is the real-time operating system that will be supported at Fermilab throughout Run II for both CDF and D0.

The CDFVME software consists of a series of Java classes to interface to generic VME boards (e.g., read and write register commands), plus templates that users can customize for their specific board. CDFVME also contains a framework in which to run a series of tests in batch mode, and to cycle the readout code. Error logging is also provided.

CDFVME uses FISION, the Fermilab supported implementation of the ANSI standard software (ANSI/VITA 25) called VISION to communicate across the VME backplane. FISION is currently supported on Motorola MV16x and PowerPC microprocessor VME CPU modules. CDFVME was jointly developed by Yale, the Online and Database Support Department of the Computing Division, and CDF.

The choice of downstream L2/3 processors in BTeV will be driven by price/performance issues. Current market trends indicate that the processors will be some type of PC processor running Linux. Linux PCs will be used for both CDF and D0 Level 3 online processor farms.

Licensing costs and various support issues are generating interest by software implementers in evaluating real time operating system alternatives to VxWorks. Real-time Linux is a possible contender for post Collider Run II data acquisition systems.

Development code for data acquisition should move in the direction of modern programming standards. Run II software has already been designed for object oriented languages such as C++ and Java. CORBA is the standard protocol for communicating between objects on remote nodes.

Chapter 11

Plan for Deployment in C0

11.1 Detector Installation

The proposed BTeV detector is not a monolithic detector like CDF or D0; it is similar in layout and construction to a traditional fixed-target detector. Most of the detector elements are individually mounted and can be separately removed from the detector for repairs. This attribute makes the initial construction and testing of the various components of the BTeV spectrometer much less dependent on the schedule of Tevatron shutdowns and maintenance periods.

The five large physical components, the vertex magnet and the four muon toroids, will be deployed first. This is important for two reasons. First, they fill most of the assembly hall during their assembly; their installation in the collision hall frees up needed space for the assembly and testing of the remaining detector elements. Second, they will form the backbone of an overhead rail and suspension system that allows the other detector elements to be inserted and removed without an overhead crane. Once installed, they are essentially passive, need never be moved, and do not affect Tevatron operations.

The beampipe in the region of the toroids and EM calorimeter does not need to be removed in order to insert or remove any detector elements. Also, both the straw tube and silicon tracking chambers clam-shell around the beampipe. The vacuum vessel containing the pixel detector can be extracted from the vertex magnet in either the upstream or downstream direction if the tracking chambers between it and the RICH tank are first rolled sideways. The crystals and photomultiplier tubes of the electromagnetic calorimeter can be serviced in place or by moving submodules from the support structure. The RICH sensors can also be serviced in place without removing the counters. This flexibility allows the RICH counters, EM calorimeter, tracking chambers, and pixel detector to be installed, tested, and repaired if necessary, in any arbitrary order in relatively brief periods of access.

11.2   Staging

Although the proposed detector is modular, installing it in the Tevatron will of course be a complicated and time consuming process. We realize test beam studies can go a long way in checking out modular components before their installation. We do envisage an extensive test beam program for the individual detector components. The trigger, on the other hand, is difficult to test in external beams. For Level 1 hardware trigger tests, it would be most helpful to install a substantial segment of the pixel detector inside the vertex magnet, in the Tevatron. Subsequently, for Level 2 hardware tests, it would also be necessary to install a portion of the forward tracking system.

Subsequently, the remaining components in one arm would be installed and detector operation could begin using relatively short periods at the end of stores or whenever luminosity became available. Installation of the second arm could occur as time and money permitted. The physics program can begin as soon as the first arm is completed.

11.3   Commissioning Plan and Goals

BTeV will be in a position to use colliding beams during the latter phases of CDF and D0 operation in Run 2B. The lab program during this period will have to carefully balance the needs of CDF and D0 for integrating luminosity with BTeV's needs for commissioning.

BTeV's goals during this period are to commission all the detector components of the first arm, and then the second. We start by getting the Level 1 and Level 2 triggers to work and proceed with the rest of the components. Our minimum goal is to have enough colliding beam exposure to carry out some of the more straight-forward measurements, those which require only a small amount of integrated luminosity. Possible first measurements would include $\sin 2\beta$ via ψK_s, B_s mixing, the study of $K^*\gamma$, and the measurement of the charm cross section in the forward direction. These will allow BTeV to debug and explore all aspects of its detector and triggers and to test its analysis programs and procedures. In addition, we believe some of these measurements will be competitive with the then-current state of the art.

We anticipate being able to check the detector out initially with collisions from a wire target and expect that this will have no impact on CDF and D0 operations. Some initial studies of the trigger may also be possible with a wire target or perhaps multiple wire targets. Low luminosity running, perhaps by using some time at the end of stores, is the next logical step and permits BTeV to check out its detector and trigger initially with one interaction per crossing. Operation at high luminosity for occasional stores to study the detector performance when the average number of interactions per crossings is high would follow. Finally, some amount of dedicated high luminosity running to accomplish the initial limited physics program above or some variant of it would be scheduled. Only this last phase of the startup is expected to have any significant impact on CDF and D0 operation. If all areas were running at once, the degradation of luminosity in CDF and D0 is expected to be no worse than 33%, during such brief periods. New ideas or circumstances may help reduce

the impact even more. BTeV believes that the value of even small amounts of running will have a huge effect on its ability to commission the detector and believes that the impact on CDF and D0 can be kept low so that their physics reach is not noticeably affected.